

**EVALUATION of the PHASE 2 EVALUATION
of the PARIS DECLARATION**

An Independent review of Strengths, Weaknesses, and Lessons



**Michael Quinn Patton
Jean Gornick
Utilization-Focused Evaluation
Saint Paul, Minnesota, USA**

July, 2011

EVALUATION of the PHASE 2 EVALUATION of the PARIS DECLARATION

An Independent review of Strengths, Weaknesses, and Lessons

Michael Quinn Patton
Jean Gornick
Utilization-Focused Evaluation
Saint Paul, Minnesota, USA

July, 2011

Contents

Executive Summary	i-iv
Introduction to the Independent Review	1
Strengths, Weaknesses, and Lessons	
Major Strengths.....	2
Major Weaknesses	7
Lessons	10
Evidence for Conclusions and In-depth Analysis	
Gathering the Perspectives of Those Involved	16
Country-Level Evaluation Process Issues.....	18
Quality of Evidence	23
Quality of Evidence Patterns and Themes	27
Final Synthesis Report Data Quality	28
Utility, Importance, Quality and Credibility of Evidence	32
Conclusion on Quality of Evidence	35
Evaluator Independence and Evaluator Collaboration.....	36
Evaluation – Stakeholder Relationships (graphic)	37
Survey Data On Collaboration and Independence	38
Concluding observations about collaboration and independence	41
Time Pressure and Meeting Deadlines	44
General vs. Particular: Aggregating and Synthesizing While Taking Context Seriously...	50
Top-Down vs. Bottoms-Up Evaluation Processes (graphic)	51
Capacity Issues	56
Impacts of Being Involved in the Paris Declaration Evaluation Process.....	60
Use	62
Expected Use of the Final Synthesis Report.....	64
The Paris Declaration Evaluation as a Model	65
Independent Audit Statement from this Evaluation of the Evaluation.....	66
Appendices (additional data from the surveys)	67

Executive Summary

Given the importance of the *Phase 2 Evaluation of the Paris Declaration*,¹ the Evaluation Management Group commissioned this independent assessment of the evaluation. It has become a standard in major high-stakes evaluations of this kind to commission an independent review to determine whether the evaluation meets generally accepted international standards of quality. That is what this report does.

Our evaluation of the Paris Declaration Evaluation included reviewing data collection instruments, templates, guidance documents, and processes; reviewing the partner country and donor evaluation reports on which the synthesis is based; directly observing two meetings of the International Reference Group (IRG) where the evidence was examined and the Final report conclusions revised accordingly; engaging IRG participants in a reflective practice session on lessons learned; surveying participants in the evaluation process and interviewing key people involved in and knowledgeable about how the evaluation was conducted. This evaluation of the evaluation includes assessing both the report's findings and the technical appendix that details how findings were generated.

We conclude that the findings, conclusions, and recommendations presented in the *Phase 2 Evaluation of the Paris Declaration* adhere closely and rigorously to the evaluation evidence collected and synthesized. Obtaining high quality evidence and thoughtfully analyzing that evidence was the constant theme of the evaluation. Both strengths and weaknesses in the evaluation are appropriately acknowledged. The comprehensive Technical Annex accurately describes data collection and analysis approaches. Partner country and donor evaluation reports, upon which the Synthesis Evaluation is based, were openly and transparently shared with the International Reference Group to allow peer review and make visible both strengths and limitations in those reports. Partner country reports were screened for adherence to quality standards with particular attention to the strength of evidence to support conclusions reached.

Those countries and donors that undertook this voluntary Phase 2 evaluation of the implementation of the Paris Declaration have engaged in systematic and in-depth reflection and evidence-based processes that make their conclusions and insights worthy of serious attention. The Synthesis Report accurately captures those evidence-based conclusions and insights. The conclusions drawn and recommendations made are appropriate for and derived from the evidence analyzed and synthesized.

This evaluation of the evaluation identifies strengths, weaknesses, and lessons for future such efforts. Each of these are discussed and documented in this report.

Major Strengths

1. Focusing attention on and inquiring deeply into the Paris Declaration and its effects.
2. Expanding knowledge of and engagement with the Paris Declaration principles.
3. Being utilization-focused throughout.
4. Keeping quality of evidence at the center of the evaluation

¹ *Phase 2 Evaluation of the Paris Declaration*

http://www.oecd.org/document/60/0,3343,en_21571361_34047972_38242748_1_1_1_1,00.html

5. Establishing the independence of the evaluation from the very beginning and scrupulously maintaining that independence throughout.
6. Making the evaluation a genuinely and authentically joint endeavor based on and adhering to the principles of the Paris Declaration.
7. Creating a standardized data gathering, assessment, and reporting framework
8. Basing the Phase 2 evaluation on the findings and lessons of Phase 1
9. Leading the evaluation effort based on vision, principles, standards, experience, competence, and commitment.
10. Assuring evaluation of the evaluation

Major Weaknesses

1. Variations in evidence quality
2. The self-selected sample of partner countries and donors
3. Variation in use of and adherence to the standardized evaluation operational matrix for partner country evaluations
4. Lack of standardized and defined criteria for rating judgments
5. Not all Phase 2 donor studies followed the revised matrix.
6. Variations in length and degree of partner country engagement.
7. Capacity variations within partner countries and greater need for capacity support than anticipated

Significant Lessons

1. Assuring evaluator independence and genuine stakeholder involvement can be mutually reinforcing in support of enhanced evaluation quality and credibility. Stakeholder involvement ensures the relevance of evaluations and evaluator independence ensures credibility. But stakeholder involvement is sometimes seen as undermining independence while processes to assure independence often limit stakeholder involvement. The Evaluation of the Paris Declaration has demonstrated how to strike the appropriate balance by creating structures and processes that provide meaningful and authentic stakeholder involvement without impinging on evaluator independence. Page 36 presents a graphic depiction of the balance that needs to be struck in evaluation-stakeholder relationships to ensure both high quality stakeholder involvement and evaluator independence that supports credibility of findings and conclusions.
2. The DAC quality standards for evaluation provide relevant, appropriate and useful guidance for conducting a complex, strategic, and principles-focused evaluation. The Final Report includes a detailed analysis of how the DAC Standards informed the evaluation.
3. The analytical, interpretative and reporting framework for assessing progress on major Paris declaration and Accra Action Agenda outcomes -- **direction, pace and distance travelled** -- proved useful in synthesizing and comparing findings, conclusions, and judgments. It is a unique and creative evaluation framework, well-suited to the challenge of synthesizing findings on progress across 21 partner country case studies
4. A comprehensive and complex international evaluation that involves multiple countries in a genuinely *joint evaluation* endeavor will need to anticipate and provide significant in-

country capacity support and development to help overcome inevitable variations in evaluation capacity. Capacity has to be built and supported for some countries throughout the evaluation.

5. An evaluation design that is “good enough,” even though not ideal, will provide credible and useful findings. The question is not whether an evaluation design is ideal, but whether it is adequate, given the constraints and challenges faced in mounting such an enormous effort. Invoking the “good enough rule,” we find that the evaluation *satisfices*. Modest claims and conclusions can be supported by relatively modest data. Strong claims and conclusions require strong supporting evidence. The Phase 2 Evaluation of Paris Declaration draws conclusions and makes recommendations appropriate to the quality of evidence generated and available.
6. Engagement of diverse international stakeholders in a global evaluation process requires careful planning, structure, execution, and facilitation. Those who would convene large and diverse stakeholder groups to advise on evaluation design, data collection, analysis, and reporting would do well to study the structure and processes of the Paris Declaration International Reference Group, especially with regard to the first lesson above, finding the appropriate balance between stakeholder involvement and ensuring evaluator independence. The joint, collaborative, and participatory nature of the evaluation meant that both bottoms-up, collaborative processes and top-down, guidance and coordination processes had to be managed. A graphic depiction of the tensions in top-down versus bottoms-up evaluation processes is presented on page 50.
7. Evaluation findings, conclusions, and recommendations are more accessible, understandable, and potentially useful for informing policy deliberations and decision-making when stated clearly, and jargon is avoided, including but not limited to evaluation jargon.
8. Evaluation findings and conclusions are more likely to invite and inform deliberation when core evaluation questions are framed as descriptive, open-ended inquiries rather than asking for simplistic judgments. It would be premature to render a summative judgment about the Paris Declaration. The evaluation design, the quality of data, and, most importantly, the complexity of development aid, do not support a summative judgment. The lesson here is to pay careful attention to the realistic and plausible purpose an evaluation can fulfill and make sure that the framing of questions are answerable and appropriate.
9. The value of an evaluation lies not just in the findings, but in the impacts that emerge among those who participate in the evaluation as they engage in the inquiry. The evaluation process increased awareness of and attention to the Paris Declaration and aid reform well beyond what would have occurred without the evaluation.
10. The joint nature and full transparency of the Evaluation of the Paris Declaration makes it a model for international evaluation. The evaluation adhered to the Paris Declaration principles even as it evaluated the implementation and results of the Paris Declaration. That is what makes it a model of international evaluation excellence.

11. Use of findings, conclusions, and recommendations can begin immediately when a high-stakes evaluation is appropriately conducted, deadlines are met, and results are disseminated widely and immediately. The Secretariat, the Evaluation Management Group, the International Reference Group, and the Core Evaluation Team attended to use from the beginning and throughout all aspects of the evaluation. It is already clear that the report is garnering widespread international attention and being used in preparations for the 4th High Level Forum on Aid Effectiveness to be held at the end of 2011.
12. Evaluation of an evaluation (meta-evaluation) can, and should be undertaken during the evaluation, not just afterward. Meta-evaluations are often undertaken only after an evaluation is completed, as a post hoc, retrospective activity. This evaluation of the evaluation was commissioned halfway through the evaluation process in time for us to observe the direction, pace, and distance travelled by the evaluation as it concluded. This made it possible to interview key participants, survey those involved, and document how the evaluation was actually conducted and the final report constructed, reviewed, and finalized. The commitment to and timing of the meta-evaluation has been exemplary.

An Independent Audit of the Synthesis Evaluation

The Final Report of the *Phase 2 Evaluation of the Paris Declaration* includes, as part of the preface, our independent audit statement of the Synthesis Evaluation. That statement is at the end of this report on page 65. That statement notes that prior to undertaking this review, we had no prior relationship with any members of the Evaluation Management Group, the Secretariat, or the Core Evaluation Team. We had complete and unfettered access to any and all evaluation documents and data, and to all members of the International Reference Group, the Evaluation Management Group, the Secretariat, and the Core Evaluation Team.

Both the audit statement and this full report on which the audit statement was based conclude that, in our opinion, the Final Synthesis Report can be trusted as independent, evidence-based, and adhering to international standards for quality evaluation. Notwithstanding inevitable limitations inherent in such a complex and comprehensive evaluation initiative, and fully disclosed in the report and discussed herein, the findings of the *Phase 2 Evaluation of the Paris Declaration* can be studied and used as trustworthy and credible.

EVALUATION of the EVALUATION of the PARIS DECLARATION

An Independent Review of Strengths, Weaknesses, and Lessons

Michael Quinn Patton
Jean Gornick
Utilization-Focused Evaluation
Saint Paul, Minnesota, USA

Introduction

It has become a standard in major high-stakes evaluations to commission an independent review to determine whether the evaluation meets generally accepted standards of quality, and, in so doing, to identify strengths, weaknesses, and lessons. The Phase 2 Evaluation of the Paris Declaration certainly qualifies as a high-stakes evaluation. As the Report states:

The Evaluation is important both for accountability and to point the way for future improvements. The underlying stakes are huge: better lives for billions of people (reflected in the approaching Millennium Development Goals for 2015); hundreds of billions of dollars expended; vital international relationships; and growing demands to see results from development aid. (p. xii)

The scope of the Evaluation is immense. It has involved more than 50 studies in 22 partner countries, reviews in 18 donor agencies, as well as several studies on special themes. It has taken place over four years, in two phases between 2007 and 2011. The current report focuses on Phase 2, the period since the Accra High Level Forum held in September, 2008. The evaluation, officially launched in September, 2009, is targeted to the next High Level Forum on aid effectiveness to be held in Busan, Korea, in December, 2011.

Given the importance of the Evaluation of the Paris Declaration, the Management Group commissioned this independent assessment of the evaluation. Prior to undertaking this review, we had no prior relationship with any members of the Management Group or the Core Evaluation Team. We had complete and unfettered access to any and all evaluation documents and data, and to all members of the International Reference Group, the Management group, and the Core Evaluation Team. Our evaluation of the evaluation included reviewing data collection instruments, templates, and processes; reviewing the partner country and donor evaluation reports on which the synthesis is based; directly observing two meetings of the International Reference Group (IRG) where the evidence was examined and the conclusions refined and sharpened accordingly; engaging IRG participants in a reflective practice lessons learned session; surveying participants in the evaluation process and interviewing key people involved in and knowledgeable about how the evaluation was conducted. This evaluation of the evaluation includes assessing both the report's findings and the technical appendix that details how findings were generated.

We begin with a summary of major strengths, weaknesses and lessons. The remainder of the report provides the evidence for these conclusions.

STRENGTHS, WEAKNESSES, and LESSONS

Our evaluation of the Paris Declaration Evaluation included reviewing data collection instruments, templates, guidance documents, and processes; reviewing the partner country and donor evaluation reports on which the synthesis is based; directly observing two meetings of the International Reference Group (IRG) where the evidence was examined and the Final report conclusions revised accordingly; engaging IRG participants in a reflective practice session on lessons learned; surveying participants in the evaluation process and interviewing key people involved in and knowledgeable about how the evaluation was conducted. Drawing on all of these data sources, we open this evaluation of the evaluation with a summary of major strengths, weaknesses, and lessons.

Major Strengths

11. *Focusing attention on and inquiring deeply into the Paris Declaration and its effects.*
12. *Expanding knowledge of and engagement with the Paris Declaration principles.*
13. *Being utilization-focused throughout.*
14. *Keeping quality of evidence at the center of the evaluation*
15. *Establishing the independence of the evaluation from the very beginning and scrupulously maintaining that independence throughout.*
16. *Making the evaluation a genuinely and authentically joint endeavor based on and adhering to the principles of the Paris Declaration.*
17. *Creating a standardized data gathering, assessment, and reporting framework*
18. *Basing the Phase 2 evaluation on the findings and lessons of Phase 1*
19. *Leading the evaluation effort based on vision, principles, standards, experience, competence, and commitment.*
20. *Assuring evaluation of the evaluation*

Discussion of major strengths

1. *Focusing attention on and inquiring deeply into the Paris Declaration and its effects.*

For two years the evaluation engaged a dedicated, knowledgeable, and diverse group of people in thinking about, gathering data on, and interpreting the effects of the Paris Declaration and Accra Action Agenda. A comprehensive and in-depth evaluation process involves defining terms, clarifying and focusing priority questions, operationalizing key concepts, formalizing the theory of change, establishing boundaries around the inquiry, building partnerships, constructing data collection protocols, building capacity to gather and analyze data, facilitating relations, and focusing attention on what is being evaluated: the implementation of the Paris Declaration and its effects on development aid. The final product, *The Evaluation of the Paris Declaration Phase 2 Final Report*, is the culmination of all this effort and will, appropriately, be the focus of attention going forward. But behind that report, both making it possible and extending its impact, is a large number of people around the world who have been deeply engaged in thinking about these issues and bringing evaluative thinking and methods to bear on critical questions of effectiveness, efficiency, and impact. It seems fair to suggest that the large and diverse group of people engaged in the many activities and distinct studies that, together, make up the whole of the evaluation, have thought as

deeply about these issues as any group anywhere in the world. An evaluation process is not just talking about something. It means studying the thing being evaluated. At its best, as in this case, it involves deep intellectual and rigorous methodological inquiry. Thus, a major strength and accomplishment of this two-year inquiry has been the very process of focusing in-depth attention on and inquiry into the Paris Declaration.

2. Expanding knowledge of and engagement with the Paris Declaration principles.

This second strength follows directly from the first. The first strength highlights the in-depth knowledge and understanding created among those engaged in the evaluation. This second strength highlights the fact that the evaluation process increased awareness of and engagement with the Paris Declaration, as well as issues of aid effectiveness, well beyond what would have occurred without the evaluation. Policy makers, ministers, civil servants, agency directors and staff, NGOs, and others who were interviewed as part of country and donor studies became more aware of and thought more deeply about the Paris Declaration as a result of being interviewed. The evaluation, therefore, served a dissemination function even as the inquiry gathered data about the degree of knowledge about and engagement with the Paris Declaration. Members of national reference groups, country evaluation teams, and donor study groups have all reported that the evaluation contributed to greater knowledge about and, in many cases, greater commitment to and implementation of the Paris Declaration. This has come to be known in the evaluation literature as “process use,” which captures the fact that evaluation inquiries have an impact on the focus of inquiry quite apart from and well before the findings are reported.

3. Being utilization-focused throughout.

The evaluation was designed and implemented to contribute to the forthcoming 4th High Level Forum on Aid Effectiveness, which is to meet in Korea at the end of 2011. The delegates to that High Level Forum, and those preparing and supporting the delegates, are the primary intended users of the evaluation. The evaluation questions focus on the priority issues that the High Level Forum is expected to discuss. The timing of the evaluation, including very tight deadlines that were rigorously adhered to and met, were based on when the evaluation report had to be produced to contribute to the deliberations of the High Level Forum. The recommendations are differentiated for and aimed at specific intended users. Three recommendations are aimed at policymakers in partner countries; three are directed specifically to policymakers in donor countries; and five are offered to policymakers in both partner countries and donor countries and agencies. This exemplifies being utilization-focused.

4. Keeping quality of evidence at the center of the evaluation

Every aspect of the evaluation was **evidence-driven**. This is clear in the supporting documents and protocols produced, from the inception report, through the standardized country data collection matrix, to the template for country reports, and the interactions at the International Reference Group meetings. The focus on data quality and evidence stemmed in part from recognition that the evaluation questions were complex, multiple and diverse methods would be needed, data collection challenges would be enormous, timelines would be extremely tight, and triangulation would be essential. While the evaluation matrix required rendering judgments about the extent to which progress was being made on various outcomes, the constant message was that those judgments had to be supported by evidence. In the end, the quality of evidence varied across questions and

countries. The Final Report is open and transparent in acknowledging those variations in quality of evidence and this evaluation of the evaluation discusses those variations at some length. Thus, those weaknesses in the data that are reported and acknowledged are a result of capacity, resource, and time constraints, inherent measurement and methodological challenges, and the complexity and comprehensiveness of the entire evaluation initiative, but not to any lack of attention to the centrality of having high quality and credible evidence as the basis for evaluation conclusions, judgments, and recommendations.

5. Establishing the independence of the evaluation from the very beginning and scrupulously maintaining that independence throughout.

Independence is essential for credibility, and credibility is essential for use. The Paris Declaration Evaluation was established under the auspices of a free-standing and independent Secretariat. Thus, it was not subservient to or dependent on any existing organization, agency, or bureaucracy. The Secretariat has cooperated with the DAC Working Party on Aid Effectiveness (WP-EFF)², but is independent of that coalition. The governing structures, the management arrangements, the engagement of the International Reference Group of diverse stakeholders, and the terms of reference for the Core Evaluation Team were all established to assure the evaluation's independence. The processes of data collection, analysis, interpretation, and stakeholder review all maintained and supported the independence of the evaluation. The Core Evaluation Team had final autonomy and ultimate ownership in determining the evaluation findings, conclusions, and recommendations. This report discusses the issue of independence, and the evidence supporting this conclusion, at length.

6. Making the evaluation a genuinely and authentically joint endeavor based on and adhering to the principles of the Paris Declaration.

While the governing arrangements were aimed at ensuring the evaluation's independence, they were also geared toward making the evaluation a truly joint exercise. The evaluation process was participatory and consultative among partner countries, donors, and international organization participants. Country ownership by partner countries was made a priority to ensure full participation and engagement. Country representatives participated in designing the synthesis matrix. Quality control was based on peer reviews and international participant feedback. Providing sufficient support to make the evaluation a trilingual exercise -- English, French and Spanish -- was aimed at ensuring full participation in and access to all aspects of the evaluation. In-country evaluation team procurement was the responsibility of and followed partner country processes. Indeed, from the beginning, the evaluation was conceptualized, designed, and implemented to adhere to Paris Declaration principles. This was not always easy and resulted in administrative challenges and, in some cases, delays in implementation, but the evidence we gathered supports the conclusion that the evaluation was a genuine and authentic joint endeavor.

² Working Party on Aid Effectiveness:

http://www.oecd.org/document/35/0,3343,en_2649_3236398_43382307_1_1_1_1,00.html

7. *Creating a standardized data gathering, assessment, and reporting framework.*

The Final Report is a synthesis of 22 partner country evaluations (including one that only participated in Phase 1), 18 donor and multilateral development agency studies³, thematic studies, sector studies, supplementary studies, and secondary and supporting data and studies. The terms of reference for the Core Evaluation Team charged them with creating a standardized template to guide the synthesis. Early on, at the first International Reference Group meeting, the metaphor emerged of treating the implementation of Paris Declaration changes as a journey, which led to focusing the synthesis framework on three critical dimensions of that journey for evaluating the Paris Declaration: *direction of travel on each key intended outcome, and the pace and distance travelled so far*. Multiple sources of evidence were gathered to support the judgments rendered, mainly qualitative evidence but also sound quantitative data where available. Aggregating and synthesizing the massive amounts of data from case studies and secondary sources was greatly facilitated and given coherence by this simple but powerful template: **direction of travel, distance, and pace**.

Moreover, the operational matrix created jointly by country participants and the core evaluation team proved especially useful in providing coherence to the evaluation.

- In our survey of national reference group and partner country evaluation team members, 87% rated the standardized matrix as “very” or “somewhat useful.”
- In our survey of International Reference Group members, 96% of donors and international organization representatives rated the standardized matrix as very or somewhat important.

8. *Basing the Phase 2 evaluation on the findings and lessons of Phase 1*

A major and important strength of the Phase 2 evaluation is that the effort was not starting from scratch. Although the Phase 1 evaluation was considerably more modest in scope and much narrower in focus than Phase 2, the first phase provided a useful foundation for the second phase evaluation. The Phase 1 evaluation highlighted the centrality of context and emphasized that processes for improving aid effectiveness are political and not just technical. Phase 1 also provided critical methodological lessons that guided the Phase 2 design, for example, the emphasis on sound evidence for comparative analysis via a clear central framework (see strength #7 above). In short, the Phase 1 evaluation offered recommendations that guided the design of Phase 2. The continuity of leadership and key evaluation team members proved to be a huge asset in completing the much more comprehensive and complex Phase 2 evaluation as a joint initiative involving many more partner countries and donors.

³ The World Bank and European Commission were notable by their absence, a fact often commented on, negatively, by those who did participate.

9. Leading the evaluation effort based on vision, principles, standards, experience, competence, and commitment.

People matter. A substantial body of research on evaluation credibility and use points to the importance of what has been called “the personal factor,” which points to the interest, commitment, engagement, and leadership of key individuals as essential to conducting high quality evaluations. Much of evaluation focuses on procedures, protocols, measures, methods, and analytical techniques. But none of that happens without people. The Paris Declaration Secretariat brought vision and experienced leadership to this enterprise, without which it would not have happened. Assembling resources, garnering the support of key institutions, using networks to negotiate critical agreements, and finding competent people to manage and implement the evaluation are leadership functions. It was at this level of visionary leadership that key decisions were made and commitments engendered to base the Evaluation of the Paris Declaration on the principles of the Paris Declaration, to adhere throughout to the DAC standards for quality evaluations, and to take on such a complex, controversial, and challenging initiative in the face of many, many naysayers who loudly declared that this could not be done. Having studied for years the factors that distinguish high quality, useful evaluations from those that are neither, we would be remiss to not call attention to the great extent to which the Paris Declaration is an exemplar of the critical role played by committed and effective leadership at all levels of this evaluation. Being privy to some of the background stories of things that had to be negotiated, problems that had to be solved, resources that had to be found, people who had to be engaged, crises that had to be handled, and challenges that had to be met, it is clear that a successful evaluation is about much, much more than systematic procedures, rigorous methods, careful analysis, and thoughtful, evidence-based judgments. The personal factor, exemplified in effective leadership, is at the core of evaluation success.

10. Assuring evaluation of the evaluation

It has become a standard in major high-stakes evaluations of this kind to commission an independent review to determine whether the evaluation meets generally accepted international standards for quality. This evaluation of the evaluation is one example of taking that commitment seriously. But the Final Report was also peer reviewed for quality, strategic and policy relevance, and communicative power by distinguished international experts (see p. v of the Final report Preface). The emerging findings of the report were thoroughly reviewed at the 3rd International Reference Group meeting in December, 2010, and the draft final report’s findings, conclusions and recommendations were systematically reviewed by 4th International Reference Group meeting in April, 2011. The Evaluation Management Group provided feedback on the evaluation processes and report in playing a quality assurance role throughout. Evaluation of the evaluation for quality enhancement, clarity of communications, and policy relevance was built into the concluding stages of the evaluation. Seriously and systematically assuring evaluation of the evaluation, with diverse stakeholder involvement, has been one of the strengths of the Paris Declaration evaluation process.

This concludes the discussion of major strengths. (See page 2 for a summary of the 10 strengths.) The next page begins the discussion of major weakness.

Major Weaknesses

8. *Variations in evidence quality*
9. *The self-selected sample of partner countries and donors*
10. *Variation in use of and adherence to the standardized evaluation operational matrix for partner country evaluations*
11. *Lack of standardized and defined criteria for rating judgments*
12. *Not all Phase 2 donor studies followed the revised matrix.*
13. *Variations in length and degree of partner country engagement.*
14. *Capacity variations within partner countries and greater need for capacity support than anticipated*

Discussion of Major Weaknesses

1. *Variations in evidence quality*

One of the strengths noted in the preceding section was *keeping quality of evidence at the center of the evaluation* (# 4). Despite that emphasis, the quality of evidence varies across partner countries, donor studies, and different evaluation questions in the synthesis. The Final Report is open and transparent in acknowledging these variations in quality of evidence. It provides “strength of evidence ratings” by the Core Evaluation Team for 17 major intended outcomes of the Paris Declaration: none are rated very good; 9 are rated good; 6 adequate; and 2 are rated poor (Final Report, pp. 18-19). This evaluation of the evaluation reports strength of evidence ratings from partner country participants and members of the International Reference Group.

Strongest evidence ratings

- The country’s ownership over development
- Overall implementation of the principles of the Paris Declaration.
- Improvements in the efficiency of aid delivery.
- Better (more inclusive and effective) partnerships.
- Overall improvements in the management and use of aid.

Weakest evidence ratings

- Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.
- Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.
- Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.
- The implications for aid effectiveness in the future.
- Delivering and accounting for development results.
- Evolution in the mix of aid modalities.

2. The self-selected sample of partner countries and donors

The voluntary nature of participation in the evaluation resulted in 21 partner countries and 7 donors doing Phase 2 case studies. The partner country sample is certainly diverse, from very large and populous countries (e.g., Indonesia and Bangladesh) to quite small countries (e.g., Cook Islands and Samoa), and from higher income countries (e.g., Colombia and the Philippines) to low income countries (e.g., Benin and Malawi), and more stable contexts (e.g., Ghana and Vietnam) to fragile states (Afghanistan). There are 10 African countries, 7 from Asia, and only two from South America. “Two proposed country evaluations (in Kyrgyzstan and Sri Lanka) were cancelled because of political change and uncertainty. Considerable effort was made to expand the number of countries in Latin and Central America but Bolivia and Colombia were the only eventual participants” (Annex 5, p. 198, Final Report). In the end, the sample included a diverse range of partner countries that could be used to extrapolate patterns and trends but is neither representative nor generalizable in a classic statistical sense. The donor sample is even smaller. Particularly noteworthy by their importance *and absence* were The World Bank and European Commission.

It is worth noting that the sample reflects the inevitably political nature of the evaluation and the decision to adhere to Paris Declaration principles in honoring the decision of partner countries about whether to participate in the evaluation process. As one Evaluation Team member commented:

Just as partners signed up to the Paris Declaration on a voluntary basis, so they signed up for the evaluation on the same terms. If the evaluation had been a purely technocratic exercise, the Evaluation Management Group could have sent in highly qualified international teams to select a completely robust sample base. But, in fact, the sample is a reflection of the political nature of process. This is important.

3. Variation in use of and adherence to the standardized evaluation operational matrix for partner country evaluations

The Final Report acknowledges that “most of the Phase 2 evaluations – the central source of evidence – did not consistently use any standard set of data sources or always apply the comparable rating scales suggested in the Evaluation Matrix. Instead, they selected from and supplemented the range of sources and ratings proposed to capture the particularities of each country situation” (p. 7). In our survey of partner country participants, 20% of evaluation team leaders and national coordinators acknowledged that they struggled to cover the full scope of questions in the matrix. This was particularly true for those partner countries that assembled their evaluation teams late and missed the opportunity to participate in capacity-building workshops on the matrix offered by the core evaluation team.

4. Lack of standardized and defined criteria for rating judgments

The operational matrix provided scales for rating judgments. For example, each country evaluation team would be asked to assess results using a scale like Very significant change, Quite significant change, Limited change, Very limited change, or No change. Actual scales varied by questions. The most common scale used for rating progress was:

- *Substantial progress*
- *Some progress*
- *Little progress*
- *None*
- *Regression*

Each country team then had to determine its own meaning for the points on the rating scale. Furthermore, teams were instructed to apply their own weighting across one or more indicators when reaching summary judgments on progress for a core question section. This is a reasonable approach to assure context sensitivity by providing standard scale terminology but allowing each team to operationalize and interpret the scales in ways that are meaningful in their context. However, this approach invites criticism from measurement specialists that the aggregation and synthesis of scales that mean different things in different contexts essentially invalidates comparisons and is fundamentally uninterpretable. Thus, from a traditional measurement perspective this approach is a glaring weakness. On the other hand, from an inclusive and social constructivist perspective, this approach is a significant strength. These differing perspectives illustrate how divergent criteria lead to contrasting judgments about what constitutes strengths and weaknesses. *From our perspective*, the lack of standardized and defined criteria for rating judgments was a pragmatically justified approach to dealing with the realities and complexities of this multi-faceted, diverse-contexts, collaborative process.

5. Not all Phase 2 donor studies followed the revised matrix.

Phase 2 donor studies followed a general template from Phase 1 and were also offered a light additional standardized matrix for their Phase 2 studies including key "mirror questions" for the country matrix, but not all applied the additional matrix. While some partner countries struggled with the standardized partner country operational matrix, most understood its importance to the synthesis, found it useful, and attempted to follow it. In contrast, an adapted template was developed to guide donor studies, but was not applied systematically (p. 5, Final Report). The adapted donor study matrix was developed only after the Phase 2 donor studies had already received the same initial terms of reference as in Phase 1 and some had started work. The Management Group and Evaluation Team discussed and negotiated the framework for donor studies to allow comparability across all donor studies (Phase 1 and Phase 2). The donor studies were used to explain and elucidate patterns in partner country studies, but not all donors updated their Phase 1 studies so the smaller sample of Phase 2 donor participants and variability in how donor studies were conducted contributed to the donor data playing a relatively smaller role in the evaluation synthesis.

6. Variations in length and degree of partner country engagement.

One-third of the national coordinators we surveyed and over half of the evaluation team leaders reported that their country "struggled with the [national] contracting process to get the evaluation underway in a timely manner." Thus, evaluation teams formed at different times with varying degrees of training and support – and time to get the work done. Those countries that started late had less time to collect high quality data and ended up relying on a few sources and available documents. In one case, the partner country evaluation team was not contracted until November 1, 2010, a mere month before the draft country report was due to the international Core Evaluation Team.

As one Evaluation Team member commented:

This [country-controlled procurement process] was a specific, high risk case of “walking the talk” of Paris Declaration principles. It was not the only cause of delays but added a lot [to delays that did occur] and the resulting wear and tear [in getting the synthesis done]. But the alternatives would have been a centralized “Northern” managed process and/or a very much longer evaluation process, jeopardizing the relevance of results.

7. Capacity variations within partner countries and greater need for capacity support than anticipated

Different evaluation teams and team leaders had varying degrees of evaluation experience and expertise. Regional workshops were conducted to enhance the capacity of country teams to use the operational matrix and gather appropriate evidence. But some country-level evaluators needed more capacity development than the regional workshops could provide and, as noted above, not all team leaders made it to a regional workshop, in some cases because of delayed procurement and contracting processes in-country. This placed a greater responsibility on the Core Evaluation Team to provide additional, individualized capacity support, a level of engagement that had not been anticipated in the original terms of reference. While 80% of evaluation team leaders in our survey reported receiving excellent or good support and guidance from the core evaluation team, 20% reported that the guidance was only fair or poor on matters like using the matrix, collecting strong evidence, and writing the country report.

This concludes the discussion of weaknesses. See page 7 for a summary of the seven weaknesses. We turn now to lessons from the experience of evaluating Phase 2 of the Paris Declaration.

LESSONS

1. Assuring evaluator independence and genuine stakeholder involvement can be mutually reinforcing in support of enhanced evaluation quality and credibility.

Stakeholder involvement assures the relevance of evaluations and evaluator independence ensures credibility. But stakeholder involvement is sometimes seen as undermining independence while processes to assure independence often limit stakeholder involvement. The Evaluation of the Paris Declaration has demonstrated how to strike the appropriate balance by creating structures and processes that provide meaningful and authentic stakeholder involvement without impinging on evaluator independence. Clearly delimited roles, skilled facilitation, and shared commitment to high quality evidence on all sides are factors that support balance and enhance the potential for use. A degree of ongoing, mutual monitoring between the international and national governance structures and evaluation teams contributed to establishing a contextually appropriate balance between evaluator independence and stakeholder involvement, with early intervention when potential difficulties arose.

Discussion of lessons continued next page.../

2. The DAC quality standards for evaluation provide relevant, appropriate and useful guidance for conducting a complex, strategic, and principles-focused evaluation.

The DAC standards were developed to guide evaluation of projects and programs. But the Paris Declaration is not a program. It is a set of principles and political commitments. As the Final report notes:

The object of the Evaluation – an agreed set of principles and commitments to improve aid effectiveness – is not a project or programme, the more normal objects of development evaluation. In a broad sense it is more like a strategy, a domain where evaluation is beginning to be tested, but the Declaration campaign has less-clear boundaries than most strategies. Interesting common elements can also be found in the growing experience in evaluating policy influence.
(Final report, p. 3)

Despite this ambiguity about exactly what was being evaluated, an ambiguity commented on by many of the partner country and donor study participants, the DAC standards proved highly appropriate and useful. Three-fourths (76%) of country participants reported that the DAC standards were “very useful” or “somewhat useful” to the partner country evaluations. They also proved appropriate and useful for this evaluation of the evaluation. The Final Report includes a detailed analysis of how the DAC Standards informed the evaluation (see pp. 218-221).

*3. The analytical, interpretative and reporting framework for assessing progress on major Paris declaration and Accra Action Agenda outcomes -- **direction, pace and distance travelled** -- proved useful in synthesizing and comparing findings, conclusions, and judgments.*

This framework was created by the Core Evaluation Team in consultation with national coordinators, the Evaluation Management Group, and the International Reference Group. Direction, pace and distance travelled treats development as a journey, including the application of agreed changes in aid programmes. It is a unique and creative evaluation framework, well-suited to the challenge of synthesizing findings on progress across 21 partner country case studies.

4. A comprehensive and complex international evaluation that involves multiple countries in a genuinely joint evaluation endeavor will need to anticipate and provide significant in-country capacity support and development to help overcome inevitable variations in evaluation capacity.

The original evaluation terms of reference for the core evaluation team did not include substantial capacity-building and support for individual country evaluation teams. The national procurement processes were expected to yield evaluation teams with sufficient capacity to undertake the partner country evaluations. The Core Evaluation Team expected to provide regional training to evaluation team leaders. However, not all evaluation team leaders were able to participate in the regional workshops. In some cases, delays in procurement of evaluation teams meant that the evaluators had not been selected when the regional workshops were held. The standardized operational matrix was not easy to use without support and training. Other capacity weaknesses surfaced as draft country reports were submitted. Adjustments were made by the Management Group and the Core Evaluation Team to increase capacity support, but this came late, and introduced the possibility that the Core Evaluation Team might be perceived as having too much influence in the completion of

country reports. This concern, about undue influence, expressed by a few at the draft stages of review during the 3rd International Reference Group meeting in Indonesia had disappeared by the 4th International Reference Group meeting in Copenhagen, when it was clear that the partner countries owned their own evaluations and all involved understood how those evaluations were used for the synthesis. Still, concerns about inadequate capacity remained. Had the capacity needs been better anticipated and planned for, a separate capacity-building and support function might have been created separate from the core evaluation team. At any rate, a lesson from this evaluation is that capacity will have to be built and supported for some countries throughout the evaluation. In that regard we would affirm the core team's lesson about capacity support:

Where new approaches are required, capacity constraints need to be anticipated and addressed. Similarly, support to the technical aspects of individual evaluations is also crucial to ensuring consistency and commonality of approach. Adequate resources for specific technical support are an essential part of the Evaluation design; they should not be seen as 'additional' or to be drawn on 'if/where required'. (Final Report, p. 218)

5. An evaluation design that is “good enough,” even though not ideal, will provide credible and useful findings.

Evaluation pioneer Peter H. Rossi formulated the “good enough rule” for evaluation admonishing evaluators to choose the best possible design, taking into account practicality and feasibility. This rule evokes Nobel-prize winning economist Herbert Simon's notion of satisficing (as opposed to optimizing or maximizing) in making decisions. *Satisfice* combines *satisfy* with *suffice* to indicate that adequacy is often the best one can do in real world situations. Modest claims and conclusions can be supported by relatively modest data. Strong claims and conclusions require strong supporting evidence. The Phase 2 Evaluation of Paris Declaration draws conclusions and makes recommendations appropriate to the quality of evidence generated and available.

The previous section of this report noted a few methodological and measurement weaknesses in the Evaluation of the Paris Declaration. The Final Report notes the limitations of the design and data collection. Those who seek to criticize the design and question the findings will find much to complain about and question. It is far from an ideal evaluation. The question, however, is not whether it is ideal, but whether it is adequate, given the constraints and challenges faced in mounting such an enormous effort. Invoking the “good enough rule,” we find that the evaluation satisfices. The evidence for that judgment is presented in this evaluation of the evaluation.

6. Engagement of diverse international stakeholders in a global evaluation process requires careful planning, structure, execution, and facilitation.

The International Reference Group (IRG) that advised the evaluation included participants from the 22 partner countries, donors, multilateral organizations, civil society representatives, and, chiefly, members of the OECD/DAC Network on Development Evaluation as well as members of the OECD/DAC Working Party on Aid Effectiveness. The IRG was co-chaired by representatives from Malawi and Sweden and met four times during Phase 2 to provide guidance to the evaluation process. We observed the 3rd and 4th meetings of the IRG in Indonesia and Copenhagen, respectively. The IRG members reviewed and commented on the evaluation framework and the inception report, emerging findings and the draft Final Report. Those who would convene large and diverse stakeholder groups to advise on evaluation design, data collection, analysis, and reporting

would do well to study the structure and processes of the Paris Declaration IRG, especially with regard to the first lesson above, finding the appropriate balance between stakeholder involvement and ensuring evaluator independence.

This report describes how those involved assess their involvement in the evaluation process and the results of the evaluation. On our survey of partner country participants, among those who felt they knew enough to rate the IRG's capacity to engage in the evaluation, 75% rated the capacity as "high" and 25% as "moderate." None rated the group's capacity as "poor." It is worth highlighting this lesson on careful planning, structure, execution, and facilitation of stakeholder involvement so that those who would undertake such processes appreciate and understand that stakeholder engagement can be difficult, challenging, expensive, controversial, conflict-laden, and undermining, and will be all those things to an extent that harms the evaluation's credibility unless the process is carefully planned and facilitated. The more diverse the group, the greater the risks. The IRG process more than satisfied. It was, in our judgment, exemplary both in design and execution.

7. Evaluation findings, conclusions, and recommendations are more accessible, understandable, and potentially useful for informing policy deliberations and decision-making when stated clearly, and jargon is avoided, including but not limited to evaluation jargon.

Throughout the evaluation, the Core evaluation Team was committed to avoiding jargon and acronyms, and the clarity of the Final Report is a tribute to their commitment in this regard. The Final Report provides conclusions and recommendations to inform the deliberations of the 4th High Forum on Aid Effectiveness and, in so doing, avoids the formative and summative jargon of evaluation, as well as other off-putting professional and methodological jargon. Much of the final meeting of the International Reference Group was devoted to clarifying and sharpening findings, conclusions, and recommendations to enhance communications and impact. The most significant difference between the review draft of the final synthesis and the Final Report was the clarity and sharpness of the conclusions and recommendations.

8. Evaluation findings and conclusions are more likely to invite and inform deliberation when core evaluation questions are framed as descriptive, open-ended inquiries rather than asking for simplistic judgments.

The original terms of reference for the evaluation called for the evaluation to be both formative and summative (see p, 194, Annex of the Final report). Indeed, one of the three primary evaluation questions was originally framed as summative, and that wording remains in the opening of Chapter 4:

“Has the implementation of Paris Declaration strengthened the contribution of aid to sustainable development results?” (Final Report, Chapter 4 opening, p. 43)

This question wording invites a simple Yes/No summative answer, even when it adds the question of how the contribution, if any, occurred. Wisely, in our view, the Executive Summary, and the Evaluation Framework and Matrix ask a more nuanced and open-ended question, one that invites deliberation and balance:

“What contributions have improvements in aid effectiveness made to sustainable development results? (Final report, p. xii).

And later: “What contributions can aid effectiveness reforms *plausibly* be judged to have made to development results?” (Chapter 1, p. 5; and p. 202, technical Annex 5)

This change in wording becomes significant in light of the evaluation’s answer:

“None of the [country] evaluations finds an immediate and direct connection between the Declaration campaign and development results achieved, but there is evidence in a solid majority of the reports that it has made at least some *plausible contributions* to better results through the pathways of the reforms traced in the previous chapter.” (p. 43)

It would be premature to render a summative judgment about the Paris Declaration. The evaluation design, the quality of data, and, most importantly, the complexity of development aid, do not support a summative judgment. The lesson here is to pay careful attention to the realistic and plausible purpose an evaluation can fulfill and make sure that the framing of questions are answerable and appropriate. Don’t raise the stakes to summative judgment when such judgment is not possible, appropriate, nor necessary.

One Core Evaluation Team member, reflecting on this issue, put the matter succinctly:

Be realistic in both the questions and the design about what is feasible, and you’re more likely to get some honest answers.

9. The value of an evaluation lies not just in the findings, but in the impacts that emerge among those who participate in the evaluation as they engage in the inquiry.

The second strength listed earlier was that the evaluation expanded knowledge of and engagement with the Paris Declaration principles and commitments. The evaluation process increased awareness of and attention to the Paris Declaration and aid reform well beyond what would have occurred without the evaluation. This added layers of impact beyond the value of the findings generated and affects calculation of the cost-benefit ratio of undertaking an evaluation, including what those involved learned about evaluation. As one donor member of the International Reference Group commented:

“The Paris Declaration evaluation had an educational value in the sense that it demonstrated to the partner countries how an international collaborative evaluation is conducted. We are confident that the participation of a large number of scholars, experts and decision makers from the developing countries has enhanced their understanding of the evaluation process, the challenges which evaluators face, and the nature, strength and limitations of the findings and recommendations.”

The impact of an evaluation process on those involved is much more likely to be substantial and sustainable when the evaluation is a joint effort -- the focus of the next lesson.

10. The joint nature and full transparency of the Evaluation of the Paris Declaration makes it a model for international evaluation.

Country participants in the evaluation were asked to rate “the extent to which the way the Paris

Declaration evaluation was conducted constitutes a model for international policy implementation evaluation.” 74% rated the evaluation model as excellent or good, affirming that they have come to see it as an international model with implications beyond this particular exemplar. Among donors and international organization participants in the IRG, 52% rated the evaluation as excellent or good as a model. In particular, our interviews affirm that the evaluation succeeded in its aspiration to be conducted as a “joint evaluation,” including the commitment to conduct the evaluation in accordance with the principles of the Paris declaration. This is what is sometimes called “walking the talk.” The evaluation adhered to the Paris Declaration principles even as it evaluated the implementation and results of the Paris Declaration. That is what makes it a model of international evaluation excellence.

Technical Annex 5 provides details on the joint evaluation and other dimensions of the approach and model. For those who want to know how the evaluation was conducted and what makes it a model, as well as lessons learned by the Core Evaluation Team, Annex 5 is invaluable. Indeed, Annex 5 is, itself, a model of how an evaluation’s technical and methodological report should be written and reported. Too few evaluations include this level of detail, disclosure, transparency, and guidance. All the documents produced as background materials and guidance documents have been posted on the DAC website and made openly available and accessible. That degree of sharing is also a model for how evaluations should be conducted and reported.

11. Use of findings, conclusions, and recommendations can begin immediately when a high-stakes evaluation is appropriately conducted, deadlines are met, and results are disseminated widely and immediately.

While this evaluation of the evaluation focuses on auditing how the evaluation was conducted and its adherence to international standards of quality, the ultimate evaluation of the evaluation will be how it is used. Hopefully, a subsequent evaluation of use will occur. But as this is written, about six weeks after the Final Report was released, it is already clear that the report is garnering widespread international attention and being used in preparations for the 4th High Level Forum on Aid Effectiveness to be held at the end of 2011. The Secretariat, the Evaluation Management Group, the International Reference Group, and the Core Evaluation Team attended to use from the beginning and throughout all aspects of the evaluation. The third strength listed earlier was that the evaluation has been utilization-focused at every stage. All those involved have planned for use. They didn’t make the all-too-common mistake of waiting until the evaluation was completed to attend to use. Indeed, the final meeting of the International Reference Group in April, 2011, was dominated by concerns about making the report useful: clarifying findings, sharpening conclusions, and ensuring that recommendations were meaningful and actionable. Early feedback from those already using the report appears to affirm that those efforts were successful and worthwhile. Further evaluation will be needed to assess actual use among participants in the 4th High Level Forum, but the early indications are the attention to utility is yielding the intended results.

12. Evaluation of an evaluation (meta-evaluation) can, and should be undertaken during the evaluation, not just afterward.

We opened this evaluation of the evaluation by noting that it has become a standard in major high-stakes evaluations to commission an independent review to determine whether the evaluation meets generally accepted standards of quality, and, in so doing, to identify strengths, weaknesses, and lessons. But meta-evaluations are often undertaken only after an evaluation is completed, as a post

hoc, retrospective activity. This evaluation of the evaluation was commissioned halfway through the evaluation process in time for us to observe the direction, pace, and distance travelled by the evaluation as it concluded. This made it possible to interview key participants, survey those involved, and document how the evaluation was actually conducted and the final report constructed, reviewed, and finalized. The commitment to and timing of the meta-evaluation, like the evaluation process itself, has been exemplary. All that remains is to assure that ultimate use is documented and evaluated.

Having reviewed major strengths, weaknesses and lessons, we turn to the details of our approach to evaluation of the Evaluation of the Paris Declaration.

GATHERING the PERSPECTIVES OF THOSE INVOLVED

As part of this evaluation of the evaluation, we conducted a survey of national reference group members and evaluation team members in the 21 Phase 2 partner countries. The survey was developed collaboratively with the Evaluation Secretariat and Core Evaluation Team following our participation in the International Reference Group meeting in Indonesia, in December 2010. At that meeting we had the opportunity to interview a number of national coordinators and partner country evaluation team leaders. We also facilitated a reflective practice session with International Reference Group participants in which they identified strengths, weaknesses, and lessons. The results of that exercise are included in this report and formed the basis for developing the survey aimed at all country-level participants. Details of the survey will be reviewed below.

We also administered a survey to members of the International reference Group who were donor and international organization representatives. That survey was administered during the 4th International Reference Group meeting in Copenhagen in late April, 2011. Those results, from 23 donor and international organization members of the International Reference Group meeting will also be presented in this report.

This evaluation of the evaluation does not include data from national and international policymakers who are ultimately the primary intended users of the *Phase 2 Evaluation of the Paris Declaration*. A future study of their judgments about and uses of the findings, conclusions, and recommendations would add an important perspective beyond those of the people directly involved in producing the findings, which is the focus of this report.

Survey of partner country participants

Members of national reference groups and partner country evaluation teams were surveyed in late March through early April, 2011. The survey was administered online with opportunities for hard copies to be completed by those with limited online access or who simply preferred a hard copy. The Minnesota Center for Survey Research at the University of Minnesota administered the survey and tabulated responses. The survey was available in English, French and Spanish. In addition, country level participants at the International Reference Group meeting in Copenhagen who had not completed the survey were given the opportunity to add their responses. In all cases, survey responses were anonymous. The only background identification question asked was whether the respondent was involved in a National Reference Group or country evaluation team function. The survey went to 209 potential respondents. The overall response rate was 38% (n=80); not everyone answered every question so the number of responses can vary by 1 to 3 per table. Of the 21

National Coordinators, 76% responded; 71% of evaluation team leaders responded. Parts of this analysis will compare the responses of National Coordinators and Evaluation Team Leaders. The survey focused on the partner country-level evaluations.

The survey included both closed-ended questions analyzed statistically and open-ended questions which we content-analyzed to determine patterns and themes, and which provide illuminative quotations included throughout this report. Sections of the survey covered the following topics:

- How well the country evaluations handled some common evaluation issues like explaining the purpose of the evaluation to key stakeholders and using appropriate methods to answer major questions.
- Perceptions and ratings on strength of evidence to support conclusions.
- Factors affecting the use of country evaluations
- Dissemination activities for and uses of the country evaluation reports
- Utility of core documents, guidance notes, input papers, and reference materials provided by the core evaluation team
- Support and quality of guidance received to conduct the country evaluation
- Capacity issues
- Overall perceptions and ratings of various aspects of the evaluation process and results

We began with the survey with questions concerning some common issues that arise in evaluation. Respondents were asked to provide their honest and confidential judgment about how well the country evaluation in which they had been involved addressed each of these issues. The table on the next page presents the aggregate results. That is followed by comparisons of the responses of National Coordinators and Evaluation Team Leaders. We begin with these results because they provide an overview of the kinds of issues involved in conducting the evaluation of the Paris Declaration *at the country level*.

The evaluations conducted in the partner countries provided the bulk of the findings on which the final synthesis report was based. As the survey results on the next page show, the experiences in the partner countries varied, as did the quality of the data gathered and the rigor of the analysis.

One national coordinator wrote in the survey: “The Country evaluation was important in assessing the results of Government policies that were implemented even before the signing of the Paris Declaration on aid effectiveness. It also provided useful insight into new initiatives by Government to strengthen aid coordination and application of more effective results management policies. The evaluation was treated as an important and serious undertaking.”

In contrast, another wrote: “There were difficulties throughout the process and, in the analysis, associating development outcomes to the PD. In our country, the widely held perception and belief is that the results realised over the years are due to a host of so many other factors that the influential or catalytic role of the PD is remote. And we found generally limited knowledge and understanding of the PD amongst stakeholders making the evaluation of its implementation a complex and difficult exercise.”

COUNTRY-LEVEL EVALUATION PROCESS ISSUES

In 2009, national coordinators from the countries participated in regional workshops facilitated by the core international evaluation team to work out the details of data collection and analysis for the priority evaluation questions of the evaluation. This led to a common evaluation template called *the matrix*. The Final Report explains in detail how the matrix was developed collaboratively with partner countries (see Figure 1, p. 198, and p. 200). The national coordinators led National reference Groups in each country in the task of selecting the evaluation team for each country. The first set of questions on our survey were aimed at getting an overall sense of how those involved experienced those country-based processes. We begin our review by reporting these results because they provide a quick overview of key issues.

Assessments of Country-Level Evaluation Strengths and Weaknesses by Country Participants

Survey results from members of National References Groups and country evaluation teams.

Participants were asked to respond with reference to the country evaluation in which each was involved. (n=77)		We did this very well (%)	We did this adequately (%)	We struggled with this (%)	We failed at this (%)	Not sure (%)
1.	Explaining the purpose of the evaluation to key stakeholder	47%	44%	6%	-	2%
2.	Managing the contracting process to get the evaluation underway in a timely manner	29	40	25	4	3
3.	Covering the full scope of questions in the matrix	28	44	24	4	-
4.	Adding country-specific questions to the matrix	24	51	16	8	1
5.	Finding and using in-country evaluators with competence to do all that the evaluation demanded	28	44	24	4	-
6.	Having a cohesive evaluation team that worked well together	38	40	19	2	-
7.	Coordination between the national coordinator, the National Reference Group, and the evaluation team	46	42	11	1	-
8.	Explaining the context and its implications in the report	30	60	10	-	-
9.	Using appropriate methods to answer major questions	28	54	13	1	4
10.	Getting access to existing data we needed	22	43	33	1	1
11.	Getting access to a broad range of stakeholders to ensure diverse and balanced perspectives	34	51	12	2	-

		We did this very well (%)	We did this adequately (%)	We struggled with this (%)	We failed at this (%)	Not sure (%)
12.	Getting a good response rate for interviews and surveys	20	49	30	1	-
13.	Getting sufficient high quality evidence to draw strong conclusions	13	53	31	2	1
14.	Triangulating data sources (using multiple methods to cross-validate findings)	22	48	20	5	5
15.	Making the sources of data for the evaluation findings transparent	28	61	10	1	-
16.	Dealing with conflicts of interest among diverse stakeholders that could affect the evaluation's credibility	11	60	25	-	4
17.	Maintaining the independence of the evaluation	40	45	14	-	1
18.	Having enough time to do a good job on the evaluation	14	39	38	8	2
19.	Getting the report done on time	31	38	29	1	1
20.	Believing that the evaluation was worth doing	48	46	4	-	2
21.	Conducting the evaluation in a way that would be truly useful in the country	32	54	11	-	2

Strengths: What the country participants report they did *very well*

1. Believing that the evaluation was worth doing 48%
2. Explaining the purpose of the evaluation to key stakeholders 47%
3. Coordination between the national coordinator, the National Reference Group, and the evaluation team 46%
4. Maintaining the independence of the evaluation 40%
5. Having a cohesive evaluation team that worked well together 38%

Sample quotes from the open-ended question on *Strengths*

- “We had a good evaluation team that worked well together. With relevant methodologies and extensive literature on aid in our country, their independent report reflects a comprehensive overview about our aid picture. And I think this is really useful and can be seen as a handbook for those who want to know about aid here.”
- “The evaluation was able to draw engagement from a broad spectrum of

stakeholders. It served as a venue to renew interest on the PD commitments, deepen engagement of other stakeholders on aid effectiveness, and create urgency to sustain important reform initiatives especially on use of country systems, managing for results and engagement with CSOs.”

- “The Team and the Office of National Coordinator coordinated the process very well and a Working Group mechanism (in addition to National Reference Group) effectively involved all stakeholders very well.”

Weaknesses: What the country participants report they struggled with or failed at.

1. Having enough time to do a good job on the evaluation	46%
2. Getting access to existing data we needed	34%
3. Getting sufficient high quality evidence to draw strong conclusions	33%
4. Getting a good response rate for interviews and surveys	31%
5. Getting the report done on time	30%

Sample quotes from the open-ended question on *Weaknesses*

- “We did struggle with obtaining information from the various government agencies because of the large volume of information scattered over various ministries. Some of the key people had either moved overseas or no longer working in key positions but somehow, those that we approached managed to locate information or directed us to other personnel who were also involved.”
- “The evaluation team struggled most with the complex, technical nature of the evaluation. It is difficult to evaluate or measure whether the PD strengthened the contribution of ODA to development results or not because it takes a long time to see such results clearly, and a result can be affected by a lot of factors. And this evaluation had difficulties due to the time constraint.”
- “Because of the time constraint as well as the number and complexity of the questions that needed to be addressed, the evaluation is not as rigorous methodologically as it should have been. A broad expertise covering macroeconomics, aid structure, behavior and dynamics, sector-specific expertise, statistics/regression, etc. are also needed to be able to respond to the complexity of the evaluation. Such kind of expertise could not be available in a team of 3. A lot of logistical requirements were also required and challenging.”

Prominent Patterns and Themes

What emerge here are patterns and themes we shall see repeated throughout this review. The country processes worked reasonably well in launching the evaluations, building momentum for the evaluation’s purpose and importance, establishing national reference groups, selecting evaluation teams, and maintaining the evaluation’s independence. The evaluation teams struggled, however, with data collection and analysis, especially given some capacity deficiencies and the time constraints under which the evaluation operated. We’ll take up these themes and examine them in more depth as we proceed. First, let’s take a look at how the national coordinators and evaluation

team leaders viewed these issues. *The people who occupied these roles have the most in-depth knowledge about how the evaluations unfolded in their countries.* The table below compares the perspectives of National Coordinators and Evaluation Team Leaders.

Rank order comparisons of evaluation processes *done very well or adequately* as rated by National Coordinators and Evaluation Team Leaders

We did this very well or adequately.

	National Coordinators (n=16)	Evaluation Team Leaders (n=15)	Difference
1. Believing that the evaluation was worth doing	100%	100%	--
2. Explaining the context and its implications in the report	100%	93%	7%
3. Conducting the evaluation in a way that would be truly useful in the country	93%	100%	7%
4. Using appropriate methods to answer major questions	94%	93%	1%
5. Explaining the purpose of the evaluation to key stakeholders	100%	80%	20%
6. Coordination between the national coordinator, the National Reference Group, and the evaluation team	100%	73%	27%
7. Maintaining the independence of the evaluation	100%	73%	27%
8. Making the sources of data for the evaluation findings transparent	93%	80%	13%
9. Getting access to a broad range of stakeholders to ensure diverse and balanced perspectives	88%	93%	5%
10. Finding and using in-country evaluators with competence to do all that the evaluation demanded	87%	60%	27%
11. Adding country-specific questions to the matrix	81%	80%	1%
12. Getting access to existing data we needed	81%	53%	28%
13. Dealing with conflicts of interest among diverse stakeholders that could affect the evaluation's credibility	69%	87%	18%

We did this very well or adequately

	National Coordinators (n=16)	Evaluation Team Leaders (n=15)	Difference
14. Triangulating data sources (using multiple methods to cross-validate findings)	81%	67%	14%
15. Covering the full scope of questions in the matrix	81%	56%	25%
16. Getting a good response rate for interviews and surveys	69%	80%	11%
17. Having enough time to do a good job on the evaluation	72%	53%	19%
18. Having a cohesive evaluation team that worked well together	69%	67%	2%
19. Getting sufficient high quality evidence to draw strong conclusions	69%	53%	16%
20. Managing the contracting process to get the evaluation underway in a timely manner	67%	47%	20%
21. Getting the report done on time	50%	73%	23%

Commentary: Overall, there is a high degree of consistency in the rank ordering of items by National Coordinators and Evaluation Team Leaders. The top four most highly ranked items only differ by one person's rating in each case. Most other items are rated and ranked similarly. The five items on which their perceptions varied the most were those where interviews indicated conflict often arose between the coordination and evaluation functions.

Item # 12 above. Getting access to existing data we needed 28% difference of perspective

Item # 6 above. Coordination between the national coordinator, the National Reference Group, and the evaluation team 27% difference

Item #7 above. Maintaining the independence of the evaluation 27% difference

Item # 10. Finding and using in-country evaluators with competence to do all that the evaluation demanded 27% difference

Item # 15. Covering the full scope of questions in the matrix 25% difference

With this overview of the country-level evaluation strengths, weaknesses, and challenges, we turn to the central issue of the evaluation: *quality of the evidence in the country evaluations.*

QUALITY of EVIDENCE

Everyone is entitled to his own opinion, but not his own facts.

Daniel Patrick Moynihan (1927–2003)
Distinguished social scientist

DAC QUALITY STANDARD 2.9: Selection of approach and methodology

The selected methodology answers the evaluation questions using credible evidence.

As this DAC standard makes explicit, a central issue in any evaluation is the **strength of evidence to support conclusions**. Many things affect strength of evidence including resources and timelines for data gathering, response rates to interviews and surveys, existing availability of relevant data, and challenges in interpreting results and drawing conclusions about contributing factors and attribution. The national participants in the Paris declaration were asked in our survey to provide their honest and confidential assessment of the strength of evidence presented in their country reports for each of the major evaluation questions listed below. These are their ratings.

Ratings of the strength of evidence presented in the country reports. (80 country respondents)		Strong evidence presented in the report	Moderate evidence presented in the report	Minimal evidence presented in the report	No evidence presented in the report
1.	Overall implementation of the principles of the Paris Declaration.	39%	54%	5%	1%
2.	Improvements in the efficiency of aid delivery.	22%	63%	14%	1%
3.	Overall improvements in the management and use of aid.	22%	60%	16%	1%
4.	The country's ownership over development.	57%	30%	11%	1%
5.	Better (more inclusive and effective) partnerships.	23%	60%	16%	1%
6.	Delivering and accounting for development results.	20%	49%	26%	4%
7.	Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.	13%	47%	32%	6%

Ratings of the strength of evidence presented in the country reports continued.../ (80 respondents)		Strong evidence presented in the report	Moderate evidence presented in the report	Minimal evidence presented in the report	No evidence presented in the report
8.	Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.	18%	34%	44%	5%
9.	Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.	14%	44%	37%	4%
10.	Evolution in the mix of aid modalities.	22%	52%	21%	4%
11.	The implications for aid effectiveness in the future.	26%	44%	21%	6%

* One not sure response omitted from some questions

Strengths and weaknesses in quality of evidence

Respondents clearly differentiated quality of evidence for different evaluation questions. They reported having obtained the strongest evidence for:

- “Country’s ownership over development,” and
- “Overall implementation of the principles of the Paris Declaration.”

They reported the weakest evidence for:

- Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion
- Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.
- Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.
- Delivering and accounting for development results

The next page highlights the data on these items.

<u>Rank order of items on which evidence was rated strong:</u>	STRONG EVIDENCE
Item # 4. The country's ownership over development	57%
Item # 1. Overall implementation of the principles of the Paris Declaration	39%
<u>Items on which evidence rated weakest</u>	MINIMAL OR NO EVIDENCE
Item # 8. Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.....	49%
Item # 9. Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.	41%
Item # 7. Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.....	38%
Item # 6. Delivering and accounting for development results.....	30%

Sample quotations on Strengths of Evidence

- “Access to people was very good indeed, especially after emailing and explaining the whole purpose of the interview. People were responsive as they feel that they need to be heard on their viewpoints as well as to understand the whole concept of the exercise. This made our work much easier and straightforward.”
- “The nature, strength and quality of evidence in the country evaluation is excellent.”
- “We worked hard to assure that quality of evidence was very much reliable and completed within timeframe.”

Sample quotations on Weaknesses of Evidence

- “Access to evidence and data from donors was difficult, hence not adequately reflected. Moreover, few donors appeared to be interested in extending full support.”
- “Evidence was poorly gathered, mainly because of poor access to information and timing. Short time for evaluation discouraged the team from placing more efforts on conducting

interviews. Thus access to information focused on what was already available or public. That was the priority for the team.”

- “Prior to designing an evaluation methodology of this complexity one should first assess data availability, especially for development results (i.e., outcomes). Some countries will have good data, others will not, so the evaluation resources and methodology should be determined as a function of what secondary data is already available. You cannot expect an evaluation to generate robust and representative primary data on development results within the available budget and a short week timeframe. Without the requisite data on development results, it would be irresponsible to comment on sustainable development or aid effectiveness in the future.”

Comparison of leaders’ perspectives on strength of country-level evidence

Below are rank order comparisons of **EVIDENCE RATED STRONG** by National Coordinators and Evaluation Team Leaders, a comparison that provides particular insight into where their judgments differ and where they are the same.

Strong evidence presented in our report

	National Coordinators (n=16)	Evaluation Team Leaders (n=15)	Difference
1. The country’s ownership over development.	80%	60%	20%
2. Overall implementation of the principles of the Paris Declaration.	38%	50%	12%
3. Evolution in the mix of aid modalities.	31%	27%	4%
4. The implications for aid effectiveness in the future.	44%	7%	37%
5. Better (more inclusive and effective) partnerships.	25%	13%	12%
6. Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.	25%	7%	18%
7. Improvements in the efficiency of aid delivery.	25%	7%	18%
8. Delivering and accounting for development results.	19%	20%	1%
9. Overall improvements in the management and use of aid.	19%	13%	6%
10. Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.	12%	7%	5%
11. Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.	19%	0%	19%

Quality of Evidence Patterns and Themes

Quality and strength of evidence varies by country and question. The national coordinators and evaluation team leaders agree that the strongest evidence was on the *country's ownership over development*.

The greatest difference in the judgments of strength of evidence between national coordinators and evaluation team leaders is on “the implications for aid effectiveness in the future.” While 44% of national coordinators perceive strong evidence on this question only 7% of evaluation team leaders consider the evidence strong.

Having reviewed the partner country reports, we can confirm that many, though not all, struggled with quality of data. Response rates on surveys were often low. The number and quality of interviews were frequently poor, and analysis of interview data was often less than systematic and rigorous. Many country reports thus relied heavily on existing data sources and reports, government and agency documents, and a few key resource people with in-depth knowledge. In short, *primary data collection was often secondary; secondary data were often primary*.

These weaknesses were due to a combination of factors: short timelines; inadequate resources; capacity limitations; and administrative and implementation challenges. As a result, several *draft* partner country reports were heavy on opinion and light on data. The core evaluation team feedback, therefore, focused on strengthening the evidence to support conclusions and judgments. The revised partner country reports reflected this mandate. Both the open and transparent peer review process of country reports at the International Reference Group meeting in Indonesia and the guidance provided by the core evaluation team strengthened substantially the evidence presented in the final partner country reports.

The constant theme at the International Reference Group (IRG) meeting was that conclusions and judgments had to be supported by evidence. The IRG process included peer reviews and open sharing of reports. It became clear to all in attendance which country reports were particularly weak. The emphasis on evidence was established and reinforced a shared norm that achieving higher quality was paramount. Core evaluation team members had coaching sessions with those country teams whose reports were weakest and/or requested such additional coaching sessions. The revised reports were considerably improved in most cases compared to earlier drafts. As a result, in the end, most partner country participants rated their reports as excellent or good:

In a concluding section of the survey, respondents were asked to rate...

the final quality of your country evaluation report.

	Excellent	Good	Fair	Poor	No opinion
80 country respondents	28%	51%	10%	8%	3%

- Nearly four-fifths (79%) rated the final quality of their country evaluation reports as excellent or good.
- Nearly one-fifth (18%) rated the final quality as fair or poor.

FINAL SYNTHESIS REPORT DATA QUALITY

The Core Evaluation Team recognized variations in data quality across questions and countries. In the Final Report, the Core Evaluation Team is careful throughout to note variations in data quality and strength of evidence. As one core evaluation team member explained: “The evidence base we applied for Synthesis had already been through a filter and findings which were weak [or] poorly evidenced were screened out.”

The Synthesis acknowledges areas where the data are weaker and emphasizes more robust findings in drawing their final conclusions. Particularly noteworthy in this regard is Table 2 in Chapter 3, *Summary of Aggregate Progress toward Intended Outcomes on Aid Effectiveness* (pp. 19-20). This table presents strength of evidence ratings for 13 intended Paris Declaration outcomes on a scale of: Very good, good, adequate, or poor. For some items, both partner countries and donors provided evidence, so strength of evidence for each was rated separately. Moreover, some objectives had more than one element The table yielded 18 ratings.

Final report *strength of evidence ratings* by the Core Evaluation Team for major intended outcomes of the Paris declaration

Very good	0	
Good	9	<ol style="list-style-type: none"> 1. Stronger national strategies & operational frameworks 2. Increased alignment of aid: Building capacity in country systems 3. Defined measures & standards of performance & accountability 4. Less duplication of efforts...& more cost-effective donor activities 5. Reformed, simplified donor policies & procedures, more collaborative 6. More predictable & multi-year commitments on aid flows 7. Delegation of authority & incentives to field staff for partnerships 8. Enhanced accountability to donors’ citizens & parliaments 9. Enhanced donor transparency for development results
Adequate	6	<ol style="list-style-type: none"> 1. Sufficient integration of global programmes & initiatives into partner countries broader development agendas 2. Stronger partner countries institutional capacities to plan, manage and implement results-driven strategies 3. Enhanced accountability to partner countries’ citizens & parliaments 4. Structural arrangements for mutual accountability 5. Among donors/agencies, less corruption & more transparency... 6. Increased alignment of aid with partner country
Poor	2	<ol style="list-style-type: none"> 1. More shared conditionalities (among donors/agencies) 2. Less corruption & more transparency among partner countries

Source: Final Report, pp. 18-19

The Core Evaluation Team categories for rating strength of evidence do not correspond exactly to the categories we used in the surveys of partner country participants and donors/international organization representatives, so a direct comparison is not possible. However, the overall direction of the distinctions between stronger and weaker evidence is consistent among all three groups: core evaluation team, partner countries, and donors//international organization representatives.

Careful differential of strength of findings

Having reviewed the evidence in the country and donor reports, and examined the use of that evidence in the Final report, **we affirm the Core Evaluation Team's ratings. They have been careful to identify and differentiate variations in strength of evidence in drawing conclusions.**

In short, the Final Synthesis Report fully meets DAC Quality Evaluation Standard 3.9:

Validity and Reliability of Information Sources.

The evaluation report describes the sources of information used (documents, respondents, administrative data, literature, etc.) in sufficient detail so that the adequacy of the information can be assessed. The evaluation report explains the selection of case studies or any samples. Limitations regarding the representativeness of the samples are identified. The evaluation cross-validates the information sources and critically assesses the validity and reliability of the data

Donors and international organization representatives ratings of strength of evidence

The next page turns to ratings of the strength of evidence in the Final Synthesis Report. These ratings are from donors and international organization representatives who participated in the 4th International Reference Group meeting in Copenhagen at the end of April, 2011.

A methodological note

It is important to understand the different time frames for and questions asked of the two distinct survey populations: (1) participants in the partner country evaluations and (2) donors and international organization representatives who were members of the International Reference Group. The survey of partner country participants was developed following the 3rd International Reference Group in Indonesia in December, 2010 and was administered shortly after partner countries submitted their final reports to the core evaluation team. That survey, as reported in previous pages, focused on the partner country evaluations.

At the 4th International Reference Group meeting in Copenhagen, donors and international organization participants and representatives were surveyed and asked to evaluate the quality of evidence in the Final Synthesis Report. While they were completing that survey, International Reference Group members from the partner countries participated in focus groups to clarify and interpret the country survey responses; some who had not completed country surveys also did so at that time.

Thus, partner country participants primarily answered questions about their own country reports. In those cases where they assessed aspects of the synthesis report (results to be presented below), they were doing so having only seen the first draft report presented in Indonesia. In contrast, the donors and international organization representatives were responding to the near-final draft of the Final Synthesis Report they had received prior to the Copenhagen meeting and which was the focus of review and feedback at the International Reference Group meeting. The next page presents those results.

**Donors and international organization representatives ratings of
strength of evidence in the Final Synthesis Report**

Ratings of the strength of evidence presented in the Final Synthesis Report. (21 respondents)		Strong evidence presented in the report	Moderate evidence presented	Minimal evidence presented in the report	No evidence or Not Sure
1.	Overall implementation of the principles of the Paris Declaration.	33%	52%	10%	5%
2.	Improvements in the efficiency of aid delivery.	24%	43%	29%	5%
3.	Overall improvements in the management and use of aid.	24%	62%	14%	--
4.	The country's ownership over development.	48%	38%	5%	10%
5.	Better (more inclusive and effective) partnerships.	24%	62%	10%	5%
6.	Delivering and accounting for development results.	5%	52%	38%	5%
7.	Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.	5%	32%	53%	10%
8.	Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.	5%	19%	62%	14%
9.	Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.	5%	44%	43%	10%
10.	Evolution in the mix of aid modalities.	10%	62%	24%	5%
11.	The implications for aid effectiveness in the future.	10%	40%	40%	10%

Sample quotations on strength of evidence in the Final report from donors and multinational organization representatives

- “The Final report is balanced – an objective attempt to analyze disparate data.”
- “The core team has synthesized the main points of individual evaluation reports well, and they have tried to add the analysis on the insufficient parts of individual reports by using

supplementary references as much as possible, with time constraints. However, I wonder whether some supplementary references are reliable or not.”

- “Balanced reporting of the evidence. Evidence based conclusions.”
- “The synthesis report is well-balanced, but there are limitations, as the report mentions itself. The situations depend on cases. It is difficult for you to get a general conclusion about the diverse world. The synthesis evaluation should have analyzed the situations by categorizing cases.”
- “The Final Report shows a good balanced approach to comments and suggestions made to the earlier draft. At this findings and reporting stage, the Core Evaluation team has seriously considered comments and feedback too, to strengthen conclusions. Transparency has been very good.”
- “Donor studies varied in quality, but were useful to help donors/agendas address bottlenecks and obstacles in implementing PD principles.”
- “Reports seem to be of mixed quality. My feeling is that some of the donor studies were not sufficiently resourced to achieve high quality and high credibility. Conclusion: We should probably have worked harder in getting good donor evaluations.”
- “The donor studies do not show the contribution to the achievement of development results. The expectation was that country reports will have strong evidence on the performance of donors. The evidence in country reports on donors is minimal.”

The appendix includes a table comparing the rank order of ratings on data quality for Country Reports and Final Synthesis Report (see page 67). Both country-based survey respondents (n=80) and donor survey respondents (n=21) ranked strength of evidence consistently.

Summary of perceptions about strongest evidence

- The country’s ownership over development
- Overall implementation of the principles of the Paris Declaration.
- Improvements in the efficiency of aid delivery.
- Better (more inclusive and effective) partnerships.
- Overall improvements in the management and use of aid.

Summary of perceptions about weakest evidence

- Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.
- Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.
- Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.

- The implications for aid effectiveness in the future.
- Delivering and accounting for development results.
- Evolution in the mix of aid modalities.

(Data source. See table on page 67 in the appendix)

Utility, Importance, Quality and Credibility of Evidence

Both the survey of partner country participants and the survey of donors and international organization representatives asked questions about utility, importance, quality, and credibility. Keep in mind that the country participants were responding shortly after submitting their country evaluations while the donors and international organization representatives were responding after receiving the near-final Final Synthesis Report draft. Thus, the timing of and context for survey responses were somewhat different, but the comparisons are still illuminating.

Quality of evidence presented in country evaluations

National Reference Group members, including the National Coordinator, and country-level evaluation team members, including the evaluation team leader, were asked to rate the “usefulness” of the evidence in the country reports.

National reference group and evaluation team members (n=80)	Very useful	Somewhat useful	A little or not at all useful	Not sure
The quality of evidence presented in the country report.	40%	50%	10%	--

Commentary: The relatively modest ratings for strength of evidence in the earlier section probably explain this modest rating for utility. Research on evaluation use has established a significant relationship between strength of evidence and utility of findings. Moderate strength of evidence leads to moderate utility.

As one respondent wrote in the survey: “Interest from stakeholders [since report completion] has not been strong. There is not much demand that the findings of the evaluation be shared. Limited credibility of the findings may have contributed to this lukewarm interest.”

Another wrote: “Overall evidence or attribution and direct linkages to Paris and Aid Effectiveness was not easy to identify; in most cases we used established data sources and assumed connections with the overall aid environment. We did as well as possible in the time available -- and a lot better than development partners who seemed to be drawing conclusions for and against Paris on very limited evidence. *It was important and useful to emphasize evidence.*”

Donors and representatives of international organizations were not in a position to rate the usefulness of country evaluations in partner countries so we asked them to assess the importance of evidence quality *in country reports* as a dimension of assessing overall Final report quality.

Donors and members of international organizations participating in the 4th International Reference Group meeting in Copenhagen were asked to rate...

the importance of the quality of evidence in the country evaluations.

Donors and international organization representatives to the International Reference Group (IRG) in Copenhagen (n= 22)	Very important dimension of the evaluation process	Somewhat important dimension of the evaluation process	Not too important to the evaluation process	Not sure
The quality of evidence presented in the country reports.	77%	14%	--	9%

This IRG group (donors and international organization representatives) were also asked to rate the quality of evidence in the country studies, the credibility of country studies, and the likely utility of country studies. These ratings were modest, as shown below.

Ratings of the overall quality, credibility, and utility of the country studies by IRG donors and international organization representatives in Copenhagen (n=22)

RATINGS of COUNTRY EVALUATIONS		In general, High	In general, Adequate	In general, Low	Mixed & variable	Not sure
1.	QUALITY of evidence in country studies	--	36%	14%	23%	27%
2.	CREDIBILITY of country studies	--	59%	4%	14%	23%
3.	UTILITY (likely) of country studies	24%	43%	--	19%	14%

In the same vein, these IRG donors and international organization representatives were asked to rate the quality, credibility, and utility of Donor Studies.

RATINGS of DONOR STUDIES (n=22)		High	Adequate	Low	Mixed	Not sure
4.	QUALITY of evidence in the DONOR STUDIES	--	59%	18%	9%	14%
5.	CREDIBILITY of the DONOR STUDIES	4%	50%	23%	9%	14%
6.	UTILITY (likely) of the DONOR STUDIES	10%	48%	19%	14%	10%

These same IRG participants (donors and international organization representatives) were asked to rate the quality, credibility, and utility of the Final Synthesis Report (n=22).

FINAL SYNTHESIS REPORT		High	Adequate	Low	Mixed	Not sure
7.	QUALITY of evidence in the FINAL SYNTHESIS REPORT	-	73%	9%	18%	--
8.	CREDIBILITY of the FINAL SYNTHESIS REPORT	18%	46%	14%	23%	--
9.	UTILITY (likely) of the FINAL SYNTHESIS REPORT	18%	36%	27%	18%	--

Commentary: The quality, credibility, and utility of the Final Synthesis Report depend, to a great extent, on the quality and credibility of the country evaluations and donor studies.

- *Country evaluation and donor study quality and credibility are necessary, but not sufficient, conditions for the quality, credibility, and, ultimately, the utility of the Final Synthesis Report.*
- It is a tribute to how the core evaluation team analyzed, presented, and synthesized the evidence in the Final Report that 73% of IRG members from donor and international organizations rated the evidence in the Final Report as adequate (though none rated it as high – item # 7 above).
- Nearly two-thirds (64%) rated the Final Report as having high or adequate credibility (item 8 above).
- Just over half (54%) rated likely utility of the Final report as high or adequate (item 9 above).

Keep in mind that these ratings were done halfway through the 4th International Reference Group meeting in Copenhagen before the Conclusions and Recommendations sections of the Final report were refined and sharpened. The feedback from IRG participants to the Core Evaluation Team in Copenhagen was that to achieve higher utility, the Conclusions and Recommendations sections of the Final Report needed to be more focused. As noted in the opening sections of this report on strengths and lessons, the Final Report reflected these concerns about enhancing utility. Indeed, a later section in this report discusses stakeholder involvement highlights and how the feedback process worked to **both** enhance utility and maintain evaluator independence. First, we present our overall conclusion about quality of evidence.

CONCLUSION on QUALITY of EVIDENCE

Quality and strength of evidence varies by question, country, and donor. The weaknesses in the evidence are fully acknowledged in the Final report. Especially noteworthy is this caveat:

[M]ost of the Phase 2 evaluations – the central source of evidence – did not consistently use any standard set of data sources or always apply the comparable rating scales suggested in the Evaluation Matrix. Instead, they selected from and supplemented the range of sources and ratings proposed to capture the particularities of each country situation. (p. 7)

What is clear, and can be stated unequivocally, is that attaining the highest possible quality of evidence was the driving force of this evaluation. Concerns about strength of evidence permeate the evaluation framing and guidance documents. Support for and feedback by the core evaluation team to country evaluators focused on supporting conclusions and judgments with evidence. The discussions at the International reference Group meetings were dominated by attention to quality of evidence.

Based on our review of the evidence, we have concluded as follows:

In our opinion, the findings and conclusions generated adhere closely and rigorously to the evaluation evidence collected. Obtaining high quality evidence and thoughtfully analyzing that evidence was the constant theme of the evaluation. Both strengths and weaknesses in the evaluation are appropriately acknowledged. The comprehensive Technical Annex accurately describes data collection and analysis approaches. Partner country and donor evaluation reports, upon which the Synthesis Evaluation is based, were openly and transparently shared with the International Reference Group to allow peer review and make visible both strengths and limitations in those reports. Partner country reports were screened for adherence to quality standards with particular attention to the strength of evidence to support conclusions reached.

Those countries and donors that undertook this voluntary evaluation of the implementation of the Paris Declaration have engaged in systematic and in-depth reflection and evidence-based processes that make their conclusions and insights worthy of serious attention. The Final Report accurately captures those evidence-based conclusions and insights.

In our opinion, the Synthesis Report can be trusted as independent, evidence-based, and adhering to international standards for quality evaluation. Notwithstanding inevitable limitations inherent in such a complex and comprehensive evaluation initiative, the overall findings can be studied and used as trustworthy and credible.

EVALUATOR INDEPENDENCE and EVALUATOR COLLABORATION

DAC Quality Evaluation Standard 3.2: *Independence of evaluators vis-à-vis stakeholders*

Evaluators are independent from the development intervention, including its policy, operations and management functions, as well as intended beneficiaries. Possible conflicts of interest are addressed openly and honestly. The evaluation team is able to work freely and without interference. It is assured of co-operation and access to all relevant information.

DAC Quality Evaluation Standard 2.5: *Stakeholder involvement*

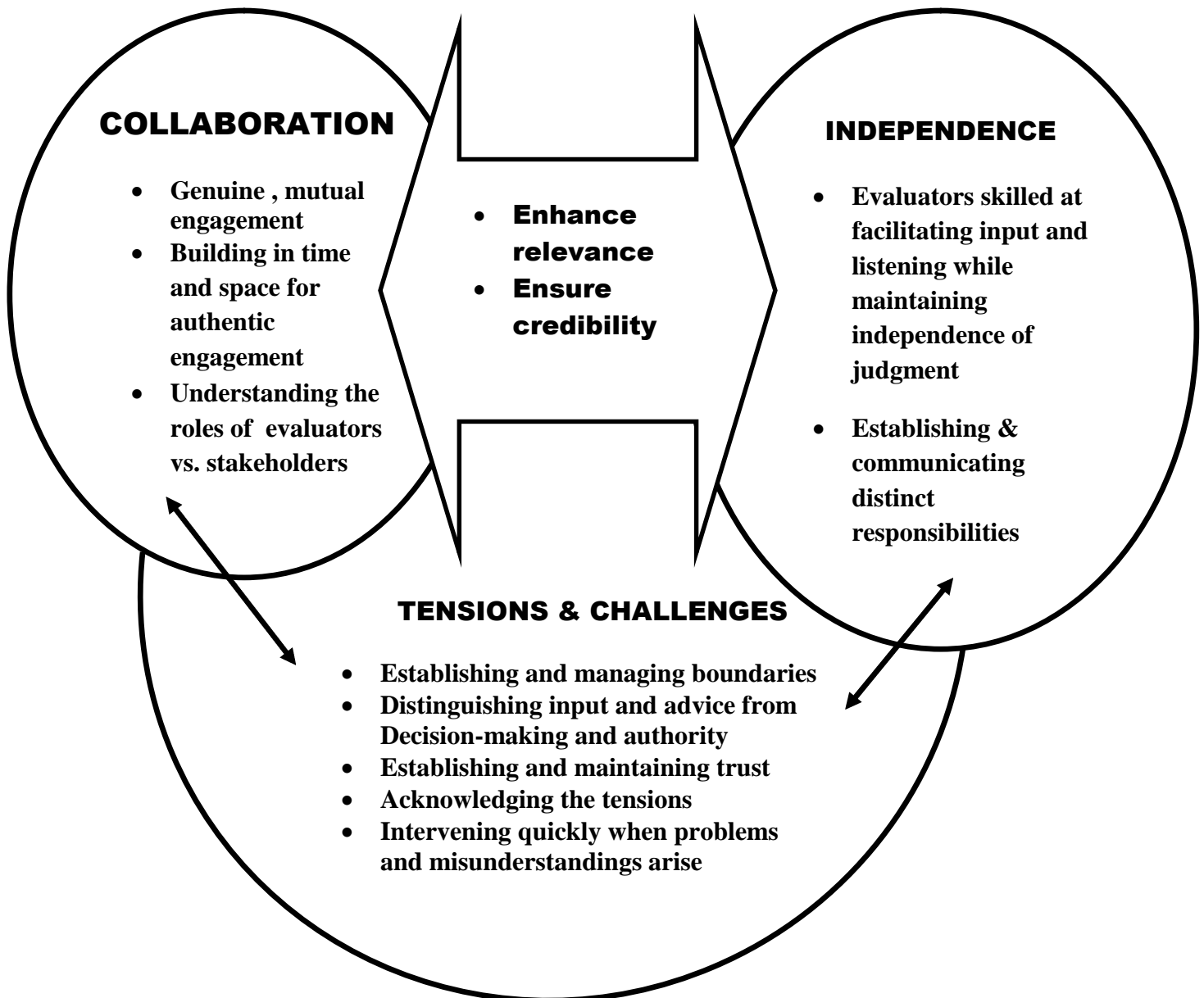
Relevant stakeholders are involved early on in the evaluation process and given the opportunity to contribute to evaluation design, including by identifying issues to be addressed and evaluation questions to be answered.

No tension more deeply permeates evaluation than the admonition that evaluators should work closely with stakeholders to enhance mutual understanding, increase relevance, and facilitate use while maintaining independence to ensure credibility. Managing this tension was at the center of the way the Evaluation of the Paris Declaration was structured, administered, and governed.

- At the country level, the national reference group in each country selected the evaluation team, based on published terms of reference and competitive processes; country evaluation were to operate independently in carrying out the evaluation.
- At the international level, the Secretariat administered the evaluation, with an Evaluation Management Group that selected the core evaluation team and provided oversight and guidance, while the Core Evaluation Team conducted the evaluation independently.
- The International Reference Group of country representatives, donor members, and international organization participants provided a stakeholder forum for engaging with the Core Evaluation Team and the evaluation's draft findings.

Thus, at every level, the evaluation was structured to separate evaluation functions from political and management functions, while also providing forums and processes for meaningful interaction. Such structures and processes are fraught with tensions and risks. Yet, in the end, the credibility and integrity of the evaluation depends on doing both well: engaging key stakeholders to ensure relevance and buy-in while maintaining independence to ensure the credibility of findings. The graphic on the next page depicts this tension.

**EVALUATOR -- STAKEHOLDER
RELATIONSHIPS**



SURVEY DATA ON COLLABORATION & INDEPENDENCE

How did those involved in the country-level evaluations view collaboration and independence? Results of three survey questions to partner country participants highlight their views. These are responses from members of the national reference groups and country evaluation teams (n=77).

Country respondents were asked to answer with reference to the country evaluation in which each was involved. (n=77)	We did this very well	We did this adequately	We struggled with this	We failed at this	Not sure
Maintaining the independence of the evaluation	40%	45%	14%	-	1%
Coordination between the national coordinator, the National Reference Group, and the evaluation team	46%	42%	11%	1%	-
Dealing with conflicts of interest among diverse stakeholders that could affect the evaluation's credibility	11%	60%	25%	-	4%

- 85% reported that they maintained independence of the evaluation adequately or very well.
- 86% viewed coordination between the national reference group and the evaluation team as adequately or very well done.
- 71% responded that they dealt with conflicts of interest adequately or very well.

Another section of the survey also asked questions related to independence and collaboration. The survey of National Reference Group (NRG) members and country evaluation team members asked respondents to rate the “**usefulness** of various factors to the country’s involvement in evaluation of the implementation of the Paris declaration.” Donors and representatives of international organizations participating in the International Reference Group (IRG) were asked in Copenhagen about the **importance** of these same elements. These different rating scales were used for the two different target survey groups because only those in each country could reasonably rate the usefulness of the country evaluations in their own countries. Donors and representatives of international organizations participating in the International Reference Group could, however, weigh in on the importance these dimensions of the evaluation process.

The results for the comparative questions relevant to independence and collaboration are presented on the next page.

These results are reported by *rank order of items rated very useful at the country-level.*

	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)
	Very useful	Very important	Somewhat useful	Somewhat important
1. Independence of the evaluation team	69%	77%	20%	14%
2. Position and responsibilities of the National Coordinator as separate from conducting the actual evaluation	54%	73%	32%	14%
3. Composition of the National Reference Group	46%	73%	31%	14%
4. Relationship between the National Reference Group and the country evaluation team	45%	59%	36%	23%
5. The process for selecting the evaluation team.	42%	82%	37%	4%

The overall consistency of these ratings between the two surveys and two groups is noteworthy:

- *Independence of the evaluation team* (item #1 above): 89% of country participants rated this as very or somewhat useful; 91% of IRG donors & international organization members rated this same item as very or somewhat important.
- *Relationship between the National Reference Group and the country evaluation team* (item # 4 above): 81% of country participants rated this relationship as very or somewhat useful; 82% of IRG donors & international organization members rated this relationship as very or somewhat important.

Open-ended comments from the surveys
(verbatim quotations from respondents)

“A particular strength of our evaluation: The independence of the evaluation team.”

“We struggled with the independence of the evaluation given the selection of the local consultants, the importance placed on their role and the constant negotiating with the National Coordinator's representative over appropriate data collection techniques, interpretation of findings and narrative prose. The result was a compromise.”

“Funding from a single donor source skewed interest and gave an impression of a lack of independence amongst some other donors.”

A weakness: “[Needed] better distinction between role of evaluation team and National Coordinator's Office in terms of evaluation neutrality and independence.”

Donor respondent: “Important lesson is independence of National Reference Groups.”

Donor respondent:” Important lesson is that the independence of country evaluation team must be ensured through a vetting process with Quality Assurance.”

Strengths, Weaknesses, and Lessons about Stakeholder Collaboration and Evaluator Independence from IRG Reflective Practice Analysis in Indonesia

Stakeholder involvement was identified as a strength of the evaluation process.

Specific strengths reported by IRG groups:

- Good to have stakeholder involvement through the national reference group from the very beginning of the evaluation
- International and national reference groups
- Inclusiveness of process – donors, recipient countries(from the start), civil society
- relative openness to CSO participation (result of CSO assertion)
- Attempt to involve as many actors as possible in the evaluation - *each having a clear role.*
- Ownership by the partner countries of the evaluation process
- Leadership and ownership of partner countries
- Participation/collaboration of key people: inputs during the whole process

Specific weaknesses reported by IRG groups:

- Difficulty in avoiding certain stakeholders influencing evaluation report formulation
- Political process
- National reference groups not always clear about their role and able to secure integrity and independence of national evaluations. (Relevant for both country level & donors/agency studies)
- Independence of evaluation teams not always secured (e.g., some evaluators are former/ex govt officials or high level govt staff, or even current in government)
- Participation of key groups in the country process weak, e.g., CSOs, donors, some ministries

- The national lead of the evaluation process through the National Reference Groups risks the independence of the evaluation process/products
- National coordinator also PD implementator and not evaluator – risks independence.
- Involvement of policy makers can be problematic: roles need to be clear and may jeopardize the integrity of the evaluation.
- Not always independence of the team judgment (pressure from the government)

Major lesson identified from IRG groups:

- Getting the governance and organizational structure right for such a complex initiative is indispensable to its success.

CONCLUDING OBSERVATIONS ABOUT COLLABORATION AND INDEPENDENCE

Attention to balancing stakeholder involvement while maintaining evaluator independence was an ever-present theme at all levels in the evaluation.

- The Secretariat, Evaluation Management Group and Core Evaluation Team all worked diligently to help National Reference Groups and country evaluation teams understand their respective roles. When there were allegations of inappropriate efforts to influence the evaluation findings or undermine evaluation team independence, Core Evaluation Team members facilitated resolutions. In a couple of cases the negotiations were intense and difficult.
- One place where the tension was most evident was in the issue of “country approvals” of the country-level evaluations. As one knowledgeable participant said: “The whole language of *country evaluations* gave the impression that these were official country documents subject to government approval. But they were the evaluation teams’ evaluations, not the country’s evaluations.”

A member of the core evaluation team explained: “[Country approval] was supposed to be a quality assurance ‘sign-off’ not approval and at every opportunity we stressed that the content of the reports was the final responsibility of the teams. I think the latter helped, but it is clear that this remained a tricky area.”

This same tension and confusion arose with some donor studies. Who owned the evaluations? Who approved them? What was subject to approval? Different understandings and different processes arose in different settings. The stakes were understood to be high because the commitment to transparency meant that evaluations would be made public.

- The Evaluation Management Group discussed some cases where independence of the evaluation was threatened and consistently reaffirmed the central importance of both stakeholder collaboration and evaluator independence. At the Synthesis level, some Evaluation Management Group members expressed strong views about certain ways of dealing with evidence, presenting conclusions, or framing recommendations, but these opinions were always followed by affirmation that the ultimate decision about how to handle any of issues related to findings, conclusions, and recommendations was the final responsibility of the Core Evaluation Team – and *it would be the Evaluation Team’s report*.

The Core Evaluation Team acted independently on inputs and evidence in ways that were clearly in line with these rules and mutual understandings.

- In the International reference Group meetings, the collaborative process allowed input on and reactions to all aspects of the Synthesis. In the December, 2010 IRG meeting in Bali, IRG members had an opportunity to react to emerging findings. The Core Evaluation Team responded systematically to how they would incorporate the feedback they received, including presenting appropriate changes in language to reflect the responses and reactions offered. This process continued and intensified at the IRG meeting in Copenhagen in April, 2011. Every chapter of the draft report was reviewed by small groups. The groups reported their reactions in plenary. The Core Evaluation Team responded to the feedback they were offered, being clear about what changes they could make based on the evidence, and always expressing appreciation for the group's input and reaffirming that, in the end, the Core Evaluation Team would determine and own the findings, conclusions, and recommendations.

A critical and revealing exchange arose near the end of the Copenhagen IRG meeting when, near the end of two days of intense interaction, a couple of IRG members pushed for one more round of review before the report was finalized to give feedback about the clarity of conclusions and recommendations. The issue was never about substance but about clear language, understandable conclusions, and communicable recommendations. At this point, the pressure on the Core Evaluation Team to complete the report on time was already quite intense. To write yet another draft and send it out for review would involve delay and threaten delivering the final report on time. Some members of the Core Evaluation Team and the Management Group expressed concern about such a delay. The tone of the demand from IRG members who wanted to see one more draft became challenging. "You tell us that you'll take our feedback into account, but now you tell us that we won't get a chance to review a new draft and react to whether you have accurately heard and incorporated our feedback."

It is at precisely at such moments that the tension and competing values between stakeholder collaboration and evaluator independence is put to the test. It was clear that the IRG process had successfully generated a strong sense of commitment to the evaluation, even a strong sense of collective ownership. This would be important as participants went home to their countries and organizations and discussed what they had experienced and expressed opinions about the authenticity of the evaluation process. Still, the Core Evaluation Team had contract obligations to meet, including a critical and inflexible deadline for delivering the Final Report.

Open discussion and negotiations ensued among key actors from the Management Group, the Secretariat, the Core Evaluation Team, and IRG members about the manageability and usefulness of a final review opportunity. A resolution emerged. The Core Evaluation Team agreed to circulate the revised Conclusions and Recommendations chapters for final feedback. Based on the interactions at the IRG, these chapters were expected to be substantially reviewed for clarity and focus. These were also the chapters that would likely receive the greatest attention going forward, including at the forthcoming 4th High level Forum. However, the revised chapters would only be available in English and there would be only one week for responses. All concerned seemed to agree that this solution was pragmatic and reasonable. And that is the process that was followed.

Members of the International Reference Group were heard. Their inputs were taken seriously. And in the end, the Core Evaluation Team took responsibility for the findings, conclusions, and recommendations based on their reading of the evidence.

From our independent and external stance, then, we can affirm the accuracy of how the Core Evaluation Team describes the collaborative process at the beginning of the report:

As the Synthesis Report and its Technical Annex show, this large and complex joint international evaluation has been a fully transparent and participatory exercise throughout. The designers and champions of the process deserve great credit for being prepared to act in the spirit of the Paris Declaration ideals. There is ample evidence that the participants have invested and benefited greatly, learning and building together. The intensive collaboration has extended from the earliest design and regional workshops all the way through to the final improvements on this Synthesis Report. (p. x)

Any joint international evaluation conducted in this collaborative manner should anticipate and be prepared to manage inevitable tensions that arise in seeking a balance between stakeholder involvement and evaluator independence. The lesson about managing this tension is so important that we made it the first lesson identified in the lessons section of this report, beginning on page 10.

Assuring evaluator independence and genuine stakeholder involvement can be mutually reinforcing in support of enhanced evaluation quality and credibility.

Stakeholder involvement assures the relevance of evaluations and evaluator independence ensures credibility. But stakeholder involvement is sometimes seen as undermining independence while processes to assure independence often limit stakeholder involvement. The Evaluation of the Paris Declaration has demonstrated how to strike the appropriate balance by creating structures and processes that provide meaningful and authentic stakeholder involvement without impinging on evaluator independence. Clearly delimited roles, skilled facilitation, and shared commitment to high quality evidence on all sides are factors that support balance and enhance the potential for use. A degree of ongoing, mutual monitoring between the international and national governance structures and evaluation teams contributed to establishing a contextually appropriate balance between evaluator independence and stakeholder involvement, with early intervention when potential difficulties arose.

The Paris declaration evaluation process was transparent, participatory, and collaborative – and the Final Report is the product of an *independent* evaluation team.

The Final report was also delivered on time which, as the story above indicates, was no small feat. We turn now to tensions around time.

TIME PRESSURE and MEETING DEADLINES

There is never enough time, unless you're serving it.

Malcolm Forbes (1919 – 1990)

Magazine publisher

DAC Quality Standard 4.1 links “timelines, relevance and use of the evaluation;”

The evaluation is delivered in time to ensure optimal use of the results.

The deadline for completing the Phase 2 Final Report was driven by the commitment to inform the deliberations of the 4th High Level Forum on Aid Effectiveness, which is to meet in Korea at the end of 2011. The agenda of that Forum will focus on reviewing implementation of the Paris Declaration and “chart the course ahead for aid effectiveness. This Final Report in combination with country evaluation reports and donor studies is expected to have wide and ongoing uses in individual countries and internationally both before and after that Forum” (p. v).

The key events of the evaluation are summarized in the timeline presented in Annex 5 (p. 198). Phase 2 of the Evaluation commenced in September 2009 with the award of the Core Team contract, following a competitive procurement process. Nine regional workshops were held in two rounds beginning in October, 2009 and continuing through August, 2010. Draft country reports were due before the meeting of the International Reference Group meeting in Indonesia in December, 2010 where emerging synthesis findings were presented. Final country and donor reports were due in January, 2011. The near-final Synthesis Report had to be produced by April, 2011, in time for review by the International Reference Group meeting in Copenhagen. Final revisions followed that meeting with the Final Report published at the beginning of June, 2011.

This was an extremely ambitious schedule. The participating countries were responsible for managing their own evaluations, including procuring independent evaluation teams. Delays in these country processes meant that country teams began their work at different times. This had consequences for the fidelity of the country evaluation processes. We’ll highlight just four consequences.

1. Uneven participation in regional workshops. The regional workshops were designed to build a genuinely joint evaluation process and build capacity. Key tasks included: defining the key principles of the methodology; shaping the evaluation questions; confirming the approaches to the methodology; working out how to operationalize the methodology; finalizing the evaluation components; and setting out guidance on methods for country and donor studies. Some countries had no participants in a regional

workshop. National coordinators were involved, but evaluation team leaders were not involved. As one member of the core evaluation team explained:

“People in the workshops were country coordinators. Not 100% representation in workshops – one or two didn’t come. There was some changeover in country coordinators. In some cases the country coordinator position had been delegated to one person but the person who makes things happen was another person - sometimes the more senior person who could make decisions. In other places, somebody comes to the meetings but not the actual country coordinator.”

The core team worked to bring late arrivals up to speed and to support evaluation teams with direct guidance, but individual coaching is no substitute for direct involvement in these participatory and capacity-building exercises. Our interviews with key informants at the International reference Group meetings and the formal lessons learned session we facilitated in Indonesia emphasized the negative consequences for those who joined the evaluation process late, especially missing the regional workshops. As one observed:

“Those who missed the regional workshops never really caught up. They especially struggled to understand the matrix and how to use it.”

2. Reduced quality of data. Those countries that started late had less time to collect high quality data and ended up relying on a few sources and available documents. In the worst case, South Africa, the evaluation team was not contracted until November 1, 2010, a mere month before the draft country report was due. The South Africa report makes clear the consequences of this late start for data collection.

A delayed appointment (beginning November 1, 2010) and therefore attenuated timelines meant that the preparation of this report had to concentrate interviews into a limited time at the end of the year when many partners were travelling, finalising their own year end reporting or preparing for leave. The availability of certain South Africa government officials was extremely limited, with, for example, senior staff from the Departments of Environment and Energy in Washington and in Cancun, Mexico for the COP 16 deliberations. Subsequent stakeholder consultations were undertaken over January and early February 2011, which partially rectified omissions and helped verify original evidence and perspective, since some could not respond due to the intensity of annual strategic planning.

3. Reduced time for analysis. Analysis and report writing are time-intensive activities. When data collection is rushed, analysis time is constrained, and report deadlines are urgent, opinions and speculation come to the fore in place of evidence-based conclusions and judgments. Some of this has to do with evaluation training and capacity, to be sure. But it also has to do with time. The core evaluation team provided feedback to all country teams about their draft reports. The most common feedback the need for more and better evidence to support conclusions and judgments. The most common

explanation we heard for inadequate evidence and insufficient evidence was lack of time.

4. Frustration with the process. The intense time pressure left some country participants frustrated with the process. One survey respondent wrote: “Because of the time constraint as well as the number and complexity of the questions that needed to be addressed, the evaluation is not as rigorous methodologically as it should have been, and that is frustrating.”

However, those that got started earlier, reported satisfaction with the process – and the results:

- “We worked together with [our partners] and the Government well and got the evaluation finished on time and it was well received by all stakeholders.”

Survey Data on Time Pressure

The survey of country participants included questions about time. These responses provide a good indication of the scope of the time pressure tension.

80 country respondents	Strongly agree	Agree	Disagree	Strongly disagree	No opinion
We had sufficient time to meet the terms of reference for the evaluation.	13%	46%	32%	8%	2%

- 59% either strongly agreed or agreed that they had sufficient time.
- 40% disagreed or strongly disagreed that they had sufficient time.

Questions on time were also included in a section of the survey that questioned what country participants felt they did well or poorly.

80 country respondents	We did this very well	We did this adequately	We struggled with this	We failed at this	Not sure
Managing the contracting process to get the evaluation underway in a timely manner	29%	40%	25%	4%	3%
Having enough time to do a good job on the evaluation.	14%	39%	38%	8%	2%
Getting the report done on time.	31%	38%	29%	1%	1%

Comparison of leaders' perspectives: National Coordinators and Evaluation Team Leaders

The National Coordinators and Evaluation Team Leaders were particularly well-placed to judge the effects of time pressure on their country-level evaluations. The table on the next page compares their perspectives.

Responded as doing this “very well” or “adequately”

	National Coordinators (n=16)	Evaluation Team Leaders (n=15)	Difference
Managing the contracting process to get the evaluation underway in a timely manner	69%	47%	22%
Having enough time to do a good job on the evaluation.	62%	53%	9%
Getting the report done on time.	50%	73%	23%

Time pressure was also a factor in the meetings of the International Reference Groups. The agenda of those meetings were tight and full, and the interactions lengthy. One donor respondent on the survey wrote: “The total time allocated to this evaluation would have had to be longer because the questions raised by a number of results would have required some time to be traced back to the original thematic and other studies. Adequate time was not available at the late stage of the 4th international reference group meeting.”

Another source of data on time pressure and its effects comes from reflective practice exercise we facilitated at the Indonesia International reference Group meeting in December, 2010. Participants working in small groups were asked to identify strengths, weaknesses, and lessons from the evaluation at that stage. Concern about time pressure was a dominant theme in the identification of weaknesses, which is understandable since this was the period of greatest time pressure: the final deadline for country reports was imminent.

Weaknesses identified in the Indonesia Reflective Practice Process with the IRG

- Lack of timeliness of some critical inputs, including personnel and funds in some cases
- Getting started late
- National procurement processes sometimes weak, slow, not harmonized with the evaluation schedule and timelines

- Delays in execution of the evaluation among those charged with Cabinet presentations
- Late starts led to short-cuts, insufficient data collection
- Tight timetable didn't allow sufficiently detailed analysis
- Limited time for validation of and meaningful consultation on synthesis conclusions

Lessons from the IRG reflection exercise

- The process is complex and takes time: Don't underestimate either the complexity or the time required
- Key to success: Good work takes time

Maintaining and Meeting Deadlines

In the end, it fell to the Evaluation Management Group to determine whether to alter the original deadlines given the variations in country contracting processes, delays in report submissions, and unevenness of data quality across different country and donor reports. Following the International Reference Group meeting in Indonesia, the Evaluation Management Group, the Secretariat, and the Core Evaluation Team met to debrief the IRG meeting and affirm next steps. We observed that meeting. A major issue was whether to adjust the deadlines given the realities of how the evaluation was unfolding and the potential consequences for report quality. Several country teams were asking for more time, especially to have time to get approval for official release of their country reports. The Core Evaluation Team asked for more time, both to work with country teams to enhance their reports as well as to complete the synthesis report. Vigorous discussion ensued.

The Evaluation Management Group (EMG) decided to maintain the official deadlines, but to informally, on a case-by-case basis, make minor adjustments to deadlines as and if needed. Some such flexibility did turn out to be necessary. The EMG also supplemented Core Team resources to cope with the additional work involved in helping countries get reports submitted. The EMG decision acknowledged the risk that a rushed country report might be of insufficient quality to be included in the final synthesis analysis. On the whole, the EMG was prepared to take that risk. The dates for the Copenhagen IRG were maintained as well as the deadline for submission of the Final Synthesis Report. *Those deadlines were met.*

It was clear from the discussion that what was driving the decision on deadlines was the DAC standard on timeliness:

The evaluation is delivered in time to ensure optimal use of the results.

Two additional points are worth making here.

First, the extremely compressed timeline for analysis and synthesis was driven by carefully calculated and essential deadlines. Completing such an extensive exercise in such a short timeframe required a team with the capacity and expertise to manage expeditiously a huge volume of data

under enormous time pressure; a team that could sift through and synthesize a massive amount of information of varying quality and clarity; a team open to feedback, even last-minute feedback about clarity and focus; and, finally, a team that could write quickly and clearly to produce coherent, understandable, and credible results. Without these skills, which the Core Evaluation Team manifested at an exemplary level, a high quality final report could not have been produced on time.

The second point is that adherence to the deadline is already paying dividends. This meta-evaluation is being completed about six weeks after release of the Final Report. We have been tracking through the Secretariat and core evaluation team leader demand for the report and presentations to key groups preparing for the 4th High Level Forum. It is clear that the findings are proving timely as the High Level Forum agenda is prepared. Any delay would have missed critical windows of opportunity.

In concluding this section about managing the pressures of time, the wisdom of former U.S. President John F. Kennedy (1960 - 1963) is relevant as a lesson evident and reaffirmed in the Paris Declaration Evaluation:

We must use time as a tool, not as a crutch.

We now move from time to context. In the new section on the next page we turn to the inevitable and inherent tension between generating general conclusions while doing justice to contextual particularities. In essence, this is the challenge of aggregating and synthesizing while taking context seriously.

GENERAL vs. PARTICULAR
Aggregating and Synthesizing
While Taking Context Seriously

The first key finding and lesson from the Phase 1 Evaluation of the Paris declaration is that:

Context is fundamental.

This understanding undergirds the “country-owned” approach adopted for the Paris Declaration Evaluation. The first of the three core questions for the evaluation is: “What are the important factors that have affected the relevance and implementation of the Paris Declaration and its potential effects on aid effectiveness and development results?” The short-hand for this question became “The PD in context.” (*Evaluation Framework and Work-plan*, p. 9)

To make these factors more explicit and prominent throughout the Evaluation, special emphasis is to be placed - through the first core question - on an unusually searching analysis of the context for the implementation of the Paris Declaration and Accra Agenda in each country where an evaluation is undertaken (*Evaluation Framework and Work-plan*, p. 11, item 32.)

So, the evaluation is a bottoms-up design with country evaluations and donor studies feeding into the synthesis findings.

But to make the synthesis manageable and facilitate aggregation of findings, the evaluation approach included building overarching, common elements:

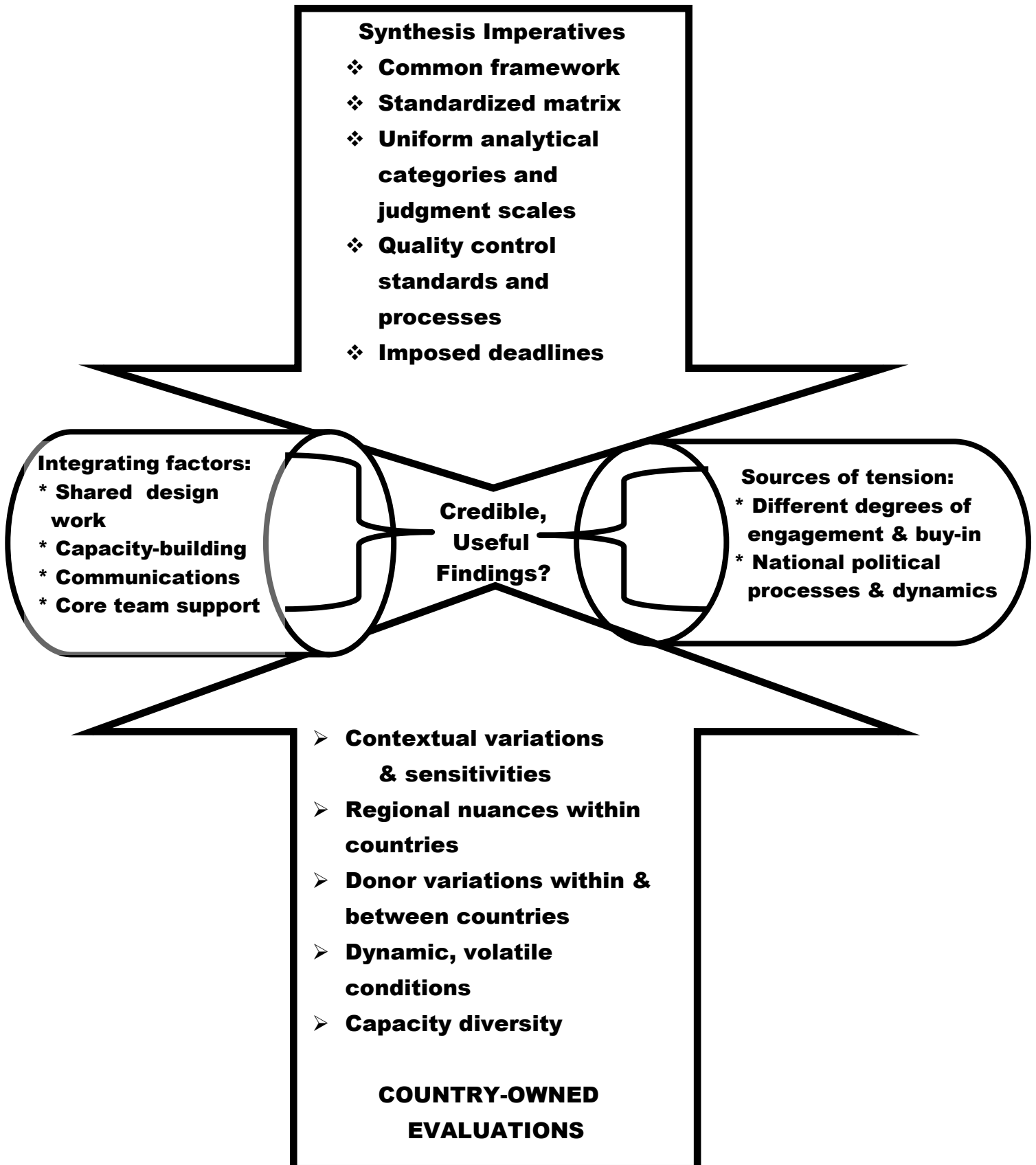
- A common evaluation framework
- Common core questions
- A standardized data collection matrix
- Standardized scales for rendering judgments
- Standardized analytical categories for describing and interpreting progress: “direction of travel, distance travelled and pace
- Common glossary of terms

So, the evaluation has a top-down, standardized framework to facilitate synthesis.

The joint, collaborative, and participatory nature of the evaluation meant that these bottoms-up and top-down processes would be managed through shared development of the details of the common framework and standardized matrix at regional workshops and ongoing capacity-building. But, as noted earlier, the uneven roll-out of the evaluation in different countries and delays in getting country evaluation teams identified and contracted meant that the actual evaluation development process involved substantially varying degrees of participation and capacity development. Extensive explanatory and guidance materials were developed to help fill this gap, but we found some confusion and tension about the relationship between country evaluations and the synthesis.

The next page provides a graphic depiction of these tensions.

TOP-DOWN VS. BOTTOMS-UP EVALUATION PROCESSES



An example of the implications of this tension is nicely illustrated by the decision about and guidance provided on country use of standardized judgment scales:

Explain rating judgements. The rating scales used have been carefully selected. The matrix will not try to pre-define the criteria for rating judgements. Each Team will need to determine their own meaning behind a particular point on the rating scale. Teams will also need to apply their own weighting across one or more indicators when reaching judgements on progress and in forming overall conclusions at the end of each core question section. (Operational Matrix for Country Evaluations, p. 1, item 7)

Examples of scales to which this guidance applies:

* Very significant change	*Significant increase	*Substantial interest/engagement
* Quite significant change	* Slight increase	*Some interest/engagement
* Limited change	* Stable	*Little interest/engagement
* Very limited change	* Slight decline	*No interest/engagement
* No change	* Significant decline	*Reduced interest/engagement

This is a reasonable approach to managing the top-down/bottom-up tension by providing standard scales but allowing each team to operationalize and interpret the scales in ways that are meaningful in their context. However, this approach invites criticism from measurement specialists that the aggregation of scales that mean different things is invalid and uninterpretable.

This also illustrates the fallibility of judgments about what constitutes strengths and weaknesses. From a hard-core measurement perspective this approach is a glaring weakness. From an inclusive and social constructivist perspective, this approach is a significant strength. From our perspective it's a pragmatically justified approach to dealing with the realities and complexities of this multi-faceted, collaborative process.

Participants' Perspectives

This tension between a country-owned evaluation process and the standardized synthesis framework first surfaced in our interviews with key informants at the International Reference Group meeting in Indonesia. One country evaluator said: "Sometimes our evaluation feels country-owned and sometimes it doesn't. We made our own data collection decisions about priorities. But the report writing is entirely dominated by the core team and their need to aggregate for the synthesis."

The reflective practice exercise we facilitated to elicit strengths, weaknesses, and lessons from the IRG participants generated several items concerning this tension. While some saw the standardized matrix as a weakness, others saw it as strength. When interpreting these findings on the next page, it is important to note once again that these data were gathered at the 3rd International Reference Group meeting when country reports were still being finalized and only preliminary, emergent synthesis findings were presented and reviewed. It was not yet clear to many participants how (or whether) a coherent whole could be constructed from the many parts, and some countries were still struggling with the looming deadline they faced to provide data on all elements of the matrix and complete their evaluations.

**Identifying *strengths, weaknesses, and lessons* from any and all aspects of the
Evaluation of the Implementation of the Paris Declaration**

Reflective practice exercise at the 3rd IRG, Indonesia, December, 2010:

61 participants at 11 tables, one French-speaking group

Strengths identified (that are relevant to this top-down vs. bottom-up tension)

- Country ownership in the process
 - Genuine attempt to make the countries part of the design
 - Countries leading their own evaluations
 - Adaptation to the situation of each country
- Comprehensive and well-structured process
- Enforced timelines/schedule; built in deadlines
- Design of the evaluation was a well-organized, multi-stage reflection process involving donors, agencies, partner countries and evaluation team leaders to reach a workable ToR and evaluation matrix.
- Same approach applied to several countries
- Evaluation is at an overarching, aggregated level which gives credibility
- Standard matrix a strength
 - Matrix provided a good comparative perspective to the evaluation framework
 - Comprehensiveness of the matrix a strength
- Involvement of the core team to ensure good quality of the component studies through advice and support
 - Review by the Core Evaluation Team: good learning process.
 - Quality of the support of the core team

Weaknesses identified (that are relevant to this tension)

- Matrix inflexible, restrictive, doesn't allow enough context at the country level
- Complicated matrix, poor internal logic
- Matrix structured to the point of being too restrictive
- The matrix itself a "straight jacket," also repetitive and duplicative.
- Language of the evaluation matrix not easily accessible
- Applicability of the matrix (weak)
- Insufficient time set aside for national team leaders to internalize the evaluation matrix and develop adequate tools (matrix not used as intended)
- Not able to capture single country uniquenesses
- Global findings didn't necessarily reflect the country level findings
- Poor alignment between donor & country studies (because donors didn't follow the matrix)
- Regional methods workshops did not have the evaluation team leaders at the methods workshops

So, some participants appreciated and praised the standardized matrix and felt they had sufficient flexibility to adapt it to the needs and context of their countries. Others found the matrix confining, inappropriate, constraining, and inflexible. These mixed results from the reflective practice exercise led us to use the survey to get a better sense of what the overall experiences and opinions were on

these matters later in the process. The survey was administered after all country reports were finalized and submitted.

Survey data on managing the tension between standardized processes and contextual sensitivities

One source of evidence on this issue is how survey respondents rated guidance and support they received from the Core Evaluation Team. National Reference Group members and evaluation team members were asked to rate the quality of the guidance and support they received from the core evaluation team in conducting the country evaluations. This support included materials provided from the core evaluation team, regional workshops offered, and consultations from core team members. Respondents were asked to make their ratings based on their own first-hand knowledge of and experience with the Phase II evaluation of the Paris Declaration.

Below is the rank order of responses for items rated Excellent or Good. (n=80)

Overall, how would you rate the guidance and support from the core evaluation team on ...

	Excellent or Good
1. how to explain the evaluation	86%
2. what the priority evaluation questions were	85
3. using the standardized matrix.	75
4. the usability and usefulness of the feedback you received from the core team on the country evaluation report.	72
5. the quality of the feedback you received from the core team on the country evaluation report.	70
6. collecting strong evidence	64
7. making the evaluation especially relevant to the country context.	63
8. the qualifications and competencies of the country's evaluators.	62
9. how the National Reference Group should operate.	61
10. writing the report	60
11. how to handle the evaluation contract process in-country.	49
12. disseminating the findings in the country.	41

For the full results of these ratings, see pages 69-70 in the Appendix.

Overall rating on support from the core evaluation team

Please respond with your country in mind. (n=80)	Strongly agree	Agree	Disagree	Strongly disagree	No opinion
We received adequate support from the international core team when needed.	22%	67%	4%	1%	6%

There are some significant differences in how National Coordinators and Evaluation Team Leaders rated guidance and support from the Core Evaluation Team. The next page compares ratings of "Excellent Guidance" because that dimension on the rating scale (Excellent, Good, Fair, Poor, No opinion) highlights variations in perspective between National Coordinators and Evaluation Team Leaders.

**Rank order of different ratings by
national coordinators and evaluation team leaders
on guidance and support from the core evaluation team**

How would you rate the guidance and support from the core evaluation team on ...	National coordinators % Excellent (n=15)	Evaluation team leaders % Excellent (n=15)	Difference
1. writing the report	46%	13%	33%
2. collecting strong evidence	40	13	27
3. using the standardized matrix	53	27	26
4. making the evaluation especially relevant to the country context	47	27	20
5. the quality of the feedback you received from the core team on the country evaluation report.	40	20	20
6. overall quality of support & guidance from the core evaluation team	47	27	20
7. what the priority evaluation questions were	36	40	4
8. disseminating the findings in the country	7	7	0
9. how to explain the evaluation	33	33	0

Commentary: The national coordinators systematically rated guidance and support from the core evaluation team higher than did evaluation team leaders, yet it was the evaluation teams that actually had to do the work of the evaluation: follow the matrix, collect data, and write the report.

We received adequate support from the international core team when needed.	Strongly agree	Agree	Disagree	Strongly disagree	No opinion
National coordinators (n=15)	53%	47%	0%	0%	0%
Evaluation team leaders (n=15)	20%	60%	13%	7%	0%

- **No national coordinators reported that support was inadequate.**
- **One-fifth (20%) of evaluation team leaders reported that they received inadequate support from the core evaluation team.**

CAPACITY ISSUES

The preceding survey items have provided an overview of how country participants perceived support and guidance from the Core Evaluation Team. This section looks at the issue of capacity-building support in more depth.

In the reflective practice exercise conducted at the Indonesia International reference Group, both strengths and weaknesses of core team support emerged. Most groups commented on the critical role of the core team in ensuring quality of the evaluations. A few participants added that for some country evaluation teams there was insufficient support from the core team for a common understanding of the key elements of the terms of reference, such as the understanding of what aid flows under the Paris Declaration principles and attribution versus contribution in drawing conclusions.

The interview data we collected from key informants on both the core evaluation team and country teams help explain these differences. The original evaluation design did not anticipate much capacity support being needed by country evaluation teams. The national reference groups were expected to find and contract with highly competent evaluators who would proceed to do their jobs. As one member of the core evaluation team explained:

Support and feedback to the countries varied from country to country. At the onset the intention was to work hard at the regional workshops to get everybody on the same page conceptually – to help them understand the whole exercise - then launch the ships.

When we launched the initial round of workshops with national coordinators, we pushed them to understand that this is not a global evaluation. This is an evaluation within each country of each country's own aid effectiveness.

After the initial round of workshops the national reference groups began their own processes – getting organized locally. Then the core team organized a second round of workshops where the emphasis was on getting the evaluation team leaders in the room so that they were on the same page, understanding things, because most teams weren't in place for the first round. As with the country coordinators, this was seen as absolutely critical to the whole process.

The first round of regional works were held in October and November, 2009 in Asia (Siem Reap), Latin America (Bogota), Anglophone Africa (Lilongwe), and Francophone Africa (Cotonou). When it became clear that many evaluation teams had not been in place for those sessions, an additional second round of workshops were added in the Pacific (Wellington, New Zealand), South Asia (Dhaka), Africa (Tunis), South East (Ha Long, Vietnam), and Latin America (Bogota and La Paz) between March and August, 2010. (For details, see Annex 5, page 200, Final report.)

But still, not all evaluation teams participated because some procurement processes continued to be delayed. This meant that the core evaluation team had to reconsider its role. As one member explained:

We realized this when we met with the Evaluation Management Group in May, 2010, having just finished the second round of regional workshops. We said to them – if you look across the 21 countries, there is a subgroup that will work well. There are enough people with experience and enough drive in the government to make it work. Our level of support would be minimal.

There is another subgroup that could probably do this pretty well, but actually they didn't have the team leader recruited yet so they didn't go through the workshop process. The national coordinator is pretty strong but he's busy and won't give much time to it. The only way it is going to roll is if we engage in some way and support that.

There was another group with willingness but capabilities might be limited. Could we play a role to help move that group forward?

We set this out to the management team and said it wasn't originally within our brief. Our original brief was to launch the ships and then connect with them when they got to the port and take the harvest. It wasn't that it had been precluded, but it wasn't articulated and was subject to need. We saw there was a real need and a potential. The synthesis process relied on it as we could only rely on what we'd get.

The core team used a variety of techniques to offer support and guidance: video conferencing, email exchanges, the web site, bi/trilateral visits, and adding staff resources to provide more teams face to face support in the southern and east Africa region. The intensity of interactions picked up as the country evaluations moved from data gathering to analysis. The focus was on helping them use the matrix and understand the need for evidence to support conclusions, but not to influence judgments.

The Technical Annex describes the core team's approach to technical support:

This technical advice role had to strike the balance between recognising the autonomy of individual teams – that is, retaining objectivity and independence – and engaging sufficiently to allow for a relatively free flow of information in terms of questions, areas of clarification etc. After some discussion, the Core Team developed a programme of support which operated mainly around the interpretation and use of the Matrix and clarification on the tools, approaches, mechanisms and methodologies of the Evaluation. (p. 209)

In reality, the matrix was challenging to implement, even for experienced evaluators. The national coordinators could not provide sufficient technical guidance in most cases. The core evaluation team tried to fulfill this capacity-building gap once it emerged, but the evidence is that they were under-resourced for the task and the time pressures of evaluators contracted late added to the difficulty. As one evaluation team leader told us:

“I felt lost when I was selected to work on the evaluation. The matrix and work involved was daunting. There was a lot to absorb, a lot to learn, a lot to do. We couldn't do it without strong core team support, but that wasn't available consistently, so we struggled.

I'm not sure of the role of the core team on feedback. We submitted a copy of our draft report informally but got no feedback until the face-to-face meeting in Bali. That was very helpful, but we needed help sooner. The time pressure was very difficult.”

We conclude that an important lesson is that in a large, complex, multi-national evaluation, some country-level evaluation teams will need substantial capacity-building and ongoing support to implement a comprehensive evaluation protocol of the kind developed for the Paris Declaration Evaluation. This suggests the need for staff and resources devoted to strengthening country team capacity with technical, methodological, and report writing support. Such additional capacity-building support might have been added to the original scope of work of the core evaluation team. However, given that the priority task of the core evaluation team was to produce the synthesis report and some concerns (few and rare, and early in the synthesis process) were raised about the core team influencing findings, an alternative approach would have been to create an independent international technical support group with one senior evaluation person per region whose job was to build capacity in and provide guidance to country teams, and help translate the matrix and synthesis evaluation framework to the country teams. Because of the importance of country evaluation capacity, this is discussed in our summary under both the sixth weakness (pp. 9-10) and the fourth lesson (p. 11).

Survey Findings on Capacity Issues

All partner country respondents were asked:

Based on your knowledge, experience, and opinion, please rate the evaluation capacity of each of the positions or groups below. *Keep in mind that your responses will be confidential.*

Partner country responses, both national coordination and evaluation teams (n=80)		High capacity to engage in evaluation (%)	Moderate capacity to engage in evaluation (%)	Low capacity to engage in evaluation (%)	Not sure about capacity (%)
1.	The country evaluation team	62%	28%	9%	1%
2.	The country evaluation team leader	63	29	7	1
3.	The National Reference Group	43	46	10	1
4.	The National Coordinator	63	32	4	1
5.	International Reference Group	65	22	-	14
6.	International core synthesis evaluation team	62	23	1	14
7.	Paris Declaration Evaluation Secretariat	63	19	-	18

The overall pattern is that two-thirds rated each group as having high capacity with the exception of the National Reference Group (item # 3) where only 43% of country-level respondents gave that group a high rating.

Given the deeper knowledge of national coordinators and evaluation team leaders, their views about capacity merit special attention. Here are their responses using the “high capacity” rating, which highlights differences in perspective.

COMPARISON of HIGH CAPACITY RATINGS

Partner country responses comparing national coordinators and evaluation team leaders on capacity ratings: HIGH CAPACITY TO ENGAGE IN EVALUATION		National Coordinators n=15 (%)	Evaluation Team Leaders n=15 (%)	Difference
1.	Country evaluation team	67%	64%	3%
2.	Country evaluation team leader	71	69	2
3.	The National Reference Group	60	20	10
4.	The National Coordinator	93	40	53
5.	International reference group	93	47	46
6.	Core synthesis evaluation team	80	53	27
7.	Evaluation Secretariat	87	54	33

Commentary: National coordinators had more confidence in evaluation teams than evaluation leaders had in national coordinators and national reference groups. In evaluating the Secretariat (item # 7), it should be noted that the evaluation teams had little direct contact; their ratings would, at best, reflect a general sense they had picked up from others, like national coordinators.

Guidance materials

Respondents were also asked to rate the usefulness of various documents provided by the Core Evaluation Team.

Rank order of materials rated VERY USEFUL

1.	Country Evaluation Operational Matrix	69%
2.	Overall Evaluation Framework	68%
3.	Generic terms of reference (TOR) for Country Evaluations	61%
4.	Country Evaluation Interview Guide	52%
5.	Phase 2 Approach Paper	50%
6.	“What is ‘aid’ in the Paris Declaration and the Evaluation?”	49%
7.	Glossary	44%
8.	“Quality Assurance and Governance Arrangements”	43%
9.	DAC Quality Standards for Development Evaluation	42%
10.	Literature review	38%
11.	“Use of Evidence in Country and Donor Evaluations”	37%
12.	“Institutional capacity and social capital”	35%
13.	Phase 1 Synthesis Report	34%
14.	“Issues of attribution and contribution in the evaluation”	34%
15.	“Contracting country evaluation teams”	32%
16.	Country Evaluation Online Discussion Forum	16%

For the full results of these ratings, see the Appendix, p. 68

IMPACTS of BEING INVOLVED in the PARIS DECLARATION EVALUATION PROCESS

An important finding from research on evaluation use in the last decade concerns the impact on those who participate in an evaluation process.

- Those directly involved, like members of National reference Groups and evaluation team members, learn an enormous amount about what they are evaluating, and they share that knowledge with others in their countries.
- Those who are interviewed, complete surveys, or provide secondary data are asked to think about the Paris Declaration. In some cases, key informants confessed to knowing little, but either in preparation for an interview or as follow-up, the evaluation stimulated them to learn more.

One national coordinator said: “Cabinet paid more attention to the Paris Declaration because we were doing this evaluation. There’s no question about it. Without the evaluation, many fewer people in government would have had any idea about the PD.”

Another responded on the survey: “The evaluation was able to draw engagement from a broad spectrum of stakeholders. It served as a venue to renew interest on the PD commitments, deepen engagement of other stakeholders on aid effectiveness, and create urgency to sustain important reform initiatives especially on use of country systems, managing for results and engagement with CSOs.”

Yet another wrote: “The evaluation enhanced general public awareness of the Paris Declaration and its application to our context.”

- Becoming more aware can lead to action. This is captured in the mantra that “what gets measured gets done.” In some cases, those involved in international aid took action to more fully implement Paris Declaration principles as a result of what they learned through the evaluation process.

One survey respondent wrote: “The evaluation process opened up a new area of consideration for Paris Declaration principles -- climate change; and we produced sub-national data which was useful.”

Another national coordinator wrote: “There have been known some significant achievements through the conduct of the evaluation on the implementation of the Paris Declaration. Stakeholders have now been capable to somehow redefine their role and function in the continuing effort to make aid more effective in development. Starting from here, a more functional building of understanding and communication among stakeholders can be realized. This example will help further strengthen the institutional requirement of having effective aid in development.”

The surveys asked about the extent to which the evaluation called attention to the Paris Declaration. The results are on the next page.

Survey item to country participants: **How useful**, in your opinion, was the way in which the evaluation called attention to the aid effectiveness agenda of the Paris Declaration in your country?

Members of partner country national reference groups and evaluation teams (n=80)	Very useful	Somewhat useful	A little bit useful	Not at all useful	Not sure
	50%	39%	9%	1%	1%

Survey item to members of the International reference Group meeting in Copenhagen (n=22): **How important**, in your opinion, was the way in which the evaluation process called attention to the aid effectiveness agenda of the Paris Declaration in partner countries?

Responses from donors and international organizations at the International reference Group meeting in Copenhagen (n=22)	Very important part of the evaluation process	Somewhat important part of the evaluation process	Not too important to the evaluation process	Not sure
	68%	23%	4%	4%

- 89% of partner country respondents reported that the evaluation was “very useful” or “somewhat useful” in calling attention to the Paris Declaration in their country.
- 91% of donors and international organizations participating in the IRG reported that calling attention to the PD was a “very important” or “somewhat important” dimension of the evaluation process.

Those countries and donors that undertook this voluntary evaluation of the implementation of the Paris Declaration have engaged in systematic and in-depth reflection and evidence-based processes that make their conclusions and insights worthy of serious attention. Members of National Reference Groups and evaluation teams have spent a great deal of time thinking about the Paris Declaration, its implications, implementation, and impacts. Those engaged as stakeholder resources, key informants, and authorizing authorities for the evaluation have engaged in thinking about the Paris Declaration principles and commitments, and what they mean for their countries. Donors that participated have likewise deepened their understanding and consideration of the implications of the Paris Declaration principles and commitments. *It seems fair to suggest that those countries and donors who have been directly engaged in the evaluation will be particularly well-prepared to participate in the 4th High Level Forum and bring evidence and systematic analysis to bear on those deliberations.*

The evidence from studies of evaluation use also indicate that those who become involved in an evaluation, especially one with such a broad-based stakeholder involvement process as this one did, are more likely to understand and use the findings.

Beginning on the next page we present the survey data on actual and expected use.

USE

Survey item: To what extent did you conduct the evaluation in a way that would be truly useful in your country?

Members of partner country national reference groups and evaluation teams (n=80)	We did this very well	We did this adequately	We struggled with this	We failed at this	Not sure
	32%	54%	11%	-	2%

- 86% percent reported conducting the evaluation either “very well” or “adequately” in a way that would be useful in their own country.

One survey respondent wrote:

- It is the most comprehensive and up-to-date public report on the country's progress in the aid effectiveness agenda. It is also the most widely disseminated, which has increased awareness at country level on the PD and the country's reform agenda.

Another wrote:

- Recent awareness of policy makers about the country evaluation may bring qualitative change in aid delivery.

But some worried that those with political power would not attend to the findings in a serious way.

- While the evaluation is a useful exercise and has potential to influence government and donor policies, the political environment in [our country] at the moment is such that any negative views expressed against government policy are not well received.
- To make it more useful, more efforts needs to done in the side of the Government and also the Development Partners. The effort is not there at present.
- In addition to the PD evaluation results, the policy makers are presented with a lot of useful information for use in decision making, but experience has shown that little action is taken to effect change for the better due to several reasons – the most notable one being lack of political will by the authorities.

Another survey item asked: Overall, how would you rate the likely use of the evaluation findings in your country?

	Excellent	Good	Fair	Poor	No opinion
Members of partner country national reference groups and evaluation teams (n=80)	20%	44%	22%	9%	5%

- **Two-thirds (66%) rated the likely use of country evaluation findings as excellent or good.**

Both International Reference Group meetings devoted time to presentation and consideration of dissemination ideas with the strong message that use of country reports was a country

responsibility, priority, and opportunity. The appendix (p. 75) includes survey data on dissemination of the country evaluation findings as of April, 2011. Survey data indicate that use of the country findings was already occurring in many countries.

Use of the Country Evaluations as of April, 2011

National Reference Group and evaluation team members were asked about use of the country evaluations through April, 2011, when the survey was conducted. These are the responses.

	National Reference Group and evaluation team members (n = 80)	Very much used	Somewhat used	A little bit used	Not at all used	Not sure
1.	So far, to what extent has the country evaluation been used by policy makers in deliberations and decisions about development aid?	6%	29%	29%	9%	27%

Expectations for Future Use of the Country Evaluations as of April, 2011

		Expect high use	Expect some use	Expect little use	Expect no use	Not sure
2.	In the future , to what extent do you expect the country evaluation to be used by policy makers in deliberations and decisions about development aid?	41%	39%	9%	1%	10%

		Expect high use	Expect some use	Expect little use	Expect no use	Not sure
3.	To what extent do you expect the country evaluation to be used by policy makers in preparation for deliberations and decisions about the 4 th High Level Forum in Korea?	58%	29%	8%	-	6%

Commentary: As noted above, the expectations for use are relatively high.

- 80% expect **high use** or **some use** of country evaluations by their country's policy makers in deliberations and decisions about development aid.
- 87% expect **high use** or **some use** of country evaluations by policy makers in preparation for deliberations and decisions about the 4th High Level Forum in Korea.

EXPECTED USE of the FINAL SYNTHESIS REPORT

Below are data on expected use of the Final Synthesis report from both the partner country survey and the survey of donors and international organization representatives. It is important to note again that survey responses from partner countries came shortly after they completed and submitted their country evaluations but before they had received the near-final draft of the synthesis report; they had only seen the emerging synthesis findings presented in Indonesia in December, 2010. Survey responses from donors and representatives of international organizations were gathered during the 4th International Reference Group in Copenhagen meeting while the near-final draft of the synthesis was being discussed.

The survey data indicate that expectations were fairly high for use of the Final Report.

Donors and International organization representatives in the IRG (n=23)	Excellent	Good	Fair	Poor
The likely usefulness of the international synthesis report to the international development community.	9%	54%	32%	4%
The likely importance of the international synthesis report at the 4th High Level Forum in Korea.	13%	56%	22%	9%

National Reference Group and evaluation team members (n = 80)	Excellent	Good	Fair	Poor	No opinion
Overall, how would you rate ...	(%)	(%)	(%)	(%)	(%)
the likely usefulness of the international synthesis report to the international development community.	29	44	13	1	13
the likely importance of the international synthesis report at the 4th High Level Forum in Korea.	30	49	12	1	9

- Nearly two-thirds (63%) of IRG donors and international organization representatives and nearly three-fourths (73%) of country participants rate as “Excellent” or “Good” the *likely usefulness of the international synthesis report to the international development community*.
- 69% of IRG donors and international organization representatives and 79% of country participants rate as “Excellent” or “Good” *the likely importance of the international synthesis Final Report at the 4th High Level Forum in Korea*.

Given these relatively high expectations of importance and use, the final section of this evaluation of evaluation examines whether the Evaluation of the Paris Declaration constitutes a model for international policy evaluation.

THE PARIS DECLARATION EVALUATION AS A *MODEL*

In our interviews with key informants we often heard that the Evaluation of the Paris Declaration should be viewed as a model for how complex global initiatives should be evaluated, so we included a question about this view on the survey.

Survey item: Overall, how would you rate the extent to which the way the Paris Declaration evaluation was conducted constitutes **a model for international policy implementation evaluation?**

	Excellent	Good	Fair	Poor	No opinion
Donors and International organization representatives in the IRG (n=23)	26%	26%	30%	4%	14%
National Reference Group and evaluation team members (n = 80)	32%	42%	18%	--	8%

- The Paris Declaration Evaluation was rated an “Excellent” or “Good” model by 52% of donors and international organization representatives at the IRG and 74% of National Reference Group and evaluation team members surveyed.

Quotes from respondents

- “A useful process for the country....Exciting to be part of an international process and to contribute to international debate on aid.”
- “The evaluation is an important guideline for aid effectiveness.”
- “A good model that should be replicated.”
- “Regular evaluation is needed for every initiative to ensure effectiveness of the initiatives and to ensure continuity in carrying forward the reform agenda.”
- “This evaluation shows how a better partnership may be established between government and development partners. As a result, this evaluation will improve and expedite effective aid delivery.”

Conclusion

This report opened with our summary of strengths, weaknesses, and lessons (pp. 1-15). The final page of this report, the next page, is the statement included in the Final Report that summarizes our Evaluation of the Evaluation of the Paris Declaration. See the Executive Summary for a review of this report.

May, 2011

An Independent Audit of the Evaluation

Readers and users of this Evaluation Report on the Paris Declaration and Aid Effectiveness may wonder, quite naturally, whether the findings can be trusted, whether the evaluation was conducted independently, and whether the evaluation process was rigorous. Just as an independent auditor's review is essential in establishing the credibility of corporate financial information to investors, stockholders and the general public, this audit of the Synthesis Evaluation speaks to the credibility of this report for intended users, policy makers, international aid stakeholders, and the global public. Given the importance of the Evaluation of the Implementation of the Paris Declaration, the Management Group commissioned this independent assessment of the evaluation. Indeed, it has become a standard in major high-stakes evaluations of this kind to commission an independent review to determine whether the evaluation meets generally accepted international standards of quality.

Prior to undertaking this review, I had no prior relationship with any members of the Management Group or the Core Evaluation Team. My associate and I had complete and unfettered access to any and all evaluation documents and data, and to all members of the International Reference Group, the Management group, and the Core Evaluation Team. Our responsibility is to express an opinion on the quality of the Synthesis Evaluation.

Our audit included reviewing data collection instruments, templates, and processes; reviewing the partner country and donors evaluation reports on which the synthesis is based; directly observing two meetings of the International Reference Group where the evidence was examined and the conclusions revised accordingly; surveying participants in the evaluation process and interviewing key people involved in and knowledgeable about how the evaluation was conducted. The evaluation audit includes assessing both the report's findings and the technical appendix that details how findings were generated.

In our opinion, the findings and conclusions generated adhere closely and rigorously to the evaluation evidence collected. Obtaining high quality evidence and thoughtfully analyzing that evidence was the constant theme of the evaluation. Both strengths and weaknesses in the evaluation are appropriately acknowledged. The comprehensive Technical Annex accurately describes data collection and analysis approaches. Partner country and donor evaluation reports, upon which the Synthesis Evaluation is based, were openly and transparently shared with the International Reference Group to allow peer review and make visible both strengths and limitations in those reports. Partner country reports were screened for adherence to quality standards with particular attention to the strength of evidence to support conclusions reached.

Those countries and donors that undertook this voluntary evaluation of the implementation of the Paris Declaration have engaged in systematic and in-depth reflection and evidence-based processes that make their conclusions and insights worthy of serious attention. The Final Report accurately captures those evidence-based conclusions and insights.

In our opinion, the Synthesis Report can be trusted as independent, evidence-based, and adhering to international standards for quality evaluation. Notwithstanding inevitable limitations inherent in such a complex and comprehensive evaluation initiative, the findings can be studied and used as trustworthy and credible.

Michael Quinn Patton, Ph. D.
Independent Evaluator and Faculty,
The Evaluators' Institute

APPENDICES

In the pages that follow, additional, more in-depth survey data are reported that informed our conclusions on strengths, weaknesses, and lessons. These tables are presented without discussion.

Comparison of rank orders of data quality for Country Reports and Final Synthesis Report

Percentage rating evidence as STRONG	NRGs & Evaluation Teams ratings of COUNTRY REPORTS (n = 80)	Donors & International Organizations ratings of SYNTHESIS REPORT (n=21)	NRGs & Eval Teams ratings (n=80)	Donors & Orgs (n=21)
	STRONG EVIDENCE		MODERATE EVIDENCE	
1. The country's ownership over development.	57%	48%	30%	38%
2. Overall implementation of the principles of the Paris Declaration.	39%	33%	54%	52%
3. Improvements in the efficiency of aid delivery.	22%	24%	63%	43%
4. Better (more inclusive and effective) partnerships.	23%	24%	60%	62%
5. Overall improvements in the management and use of aid.	22%	24%	60%	62%
	MINIMAL OR NO EVIDENCE		MODERATE EVIDENCE	
6. Improving the prioritization of the needs of the poorest people, including women and girls, and reducing social exclusion.	49%	76%	34%	19%
7. Whether implementation of the Paris Declaration has generally strengthened the contribution of aid to sustainable development.	38%	66%	47%	29%
8. Sustainable increases in institutional capacities and social capital at all levels to respond to development challenges.	41%	53%	44%	43%
9. The implications for aid effectiveness in the future.	27%	45%	44%	40%
10. Delivering and accounting for development results	30%	43%	49%	52%
11. Evolution in the mix of aid modalities.	25%	24%	52%	62%

Utility of core documents, guidance notes, input papers, & reference materials

The survey of National Reference Group and evaluation team members included an assessment of the utility to the country evaluation process of various guidance documents, input papers, and reference materials from the Paris Declaration Evaluation Secretariat and the core evaluation team. These materials were used in different ways and to varying degrees by evaluation teams in different countries. Respondents were asked to rate the utility of specific guidance materials based on their knowledge of and experience with the evaluation of the Paris Declaration.

Utility of core documents, guidance notes, input papers, & reference materials		Very useful	Somewhat useful	Not used	Don't know
National Reference Group and evaluation team members (n=80)		to our evaluation process (%)	to our evaluation process (%)	in our evaluation process (%)	if it was used (%)
1.	Phase 1 Synthesis Report	34%	52%	6%	8%
2.	Phase 2 Approach Paper	50	38	5	6
3.	Overall Evaluation Framework	68	32	--	--
4.	Country Evaluation Operational Matrix	69	30	1	--
5.	Country Evaluation Interview Guide	52	40	8	--
6.	Country Evaluation Online Discussion Forum	16	42	34	8
7.	"Contracting country evaluation teams"	32	42	11	15
8.	Generic terms of reference (TOR) for Country Evaluations	61	37	1	1
9.	Literature review	38	51	5	6
10.	Glossary	44	40	10	6
11.	"What is 'aid' in the Paris Declaration and the Evaluation?"	49	31	14	6
12.	"Quality Assurance and Governance Arrangements"	43	34	16	6
13.	"Use of Evidence in Country and Donor Evaluations"	37	42	15	5
14.	"Institutional capacity and social capital"	35	34	20	11
15.	"Issues of attribution and contribution in the evaluation"	34	44	14	8
16.	DAC Quality Standards for Development Evaluation	42	34	14	10

Guidance and Support from the Core Evaluation Team

National Reference Group members and evaluation team members were asked to rate the quality of the guidance and support they received from the core evaluation team in conducting the country evaluations. (n=80)

This support included materials provided from the core evaluation team, regional workshops offered, and consultations from core team members. Respondents were asked to make their ratings based on their own first-hand knowledge of and experience with the Phase II evaluation of the Paris Declaration.

Overall, how would you rate the guidance & support from the core team on ...		Excellent (%)	Good (%)	Fair (%)	Poor (%)	No opinion (%)
1.	how to explain the evaluation.	33%	53%	8%	1%	5%
2.	what the priority evaluation questions were.	32	53	10	1	4
3.	how to handle the evaluation contract process in-country.	17	32	21	7	24
4.	how the National Reference Group should operate.	23	38	18	8	13
5.	the qualifications and competencies of the country's evaluators.	26	36	17	8	14
6.	using the standardized matrix.	29	46	20	4	1
7.	collecting strong evidence.	16	48	25	8	2
8.	writing the report.	22	38	21	7	12
9.	making the evaluation especially relevant to the country context.	25	38	14	11	11
10.	disseminating the findings in the country.	14	27	23	13	23

Results continued next page.../

	Overall, how would you rate ...	Excellent (%)	Good (%)	Fair (%)	Poor (%)	No opinion (%)
11.	support and guidance from the Paris Declaration Evaluation Secretariat.	32	44	8	2	4
12.	support and guidance from the international core evaluation team.	30	49	15	1	4
13.	the quality of the feedback you received from the core team on the country evaluation report.	24	46	14	4	13
14.	the usability and usefulness of the feedback you received from the core team on the country evaluation report.	18	54	15	1	11
15.	the final quality of the country evaluation report.	28	51	10	3	8
16.	the likely use of the evaluation findings in the country.	20	44	22	9	5
17.	the likely quality of the international synthesis report based on the Bali draft and discussions.	20	43	9	--	28
18.	the likely usefulness of the international synthesis report to the international development community.	29	44	13	1	13
19.	the likely importance of the international synthesis report at the 4th High Level Forum in Korea.	30	49	12	1	9
20.	the extent to which the way the Paris Declaration evaluation was conducted constitutes a model for international policy implementation evaluation.	32	42	18	--	9

**UTILITY and IMPORTANCE RATINGS
Of VARIOUS ELEMENTS of the EVALUATION PROCESS**

The Paris declaration evaluation involved many elements and dimensions. The survey of National Reference Group (NRG) members asked respondents to rate these various elements and dimensions in terms of their “**usefulness** to the country’s involvement in evaluation of the implementation of the Paris declaration.” Donors and representatives of international organizations participating in the International Reference Group (IRG) were asked in Copenhagen about the **importance** of these same elements. So, country-level NRG members and evaluators were asked about utility of these elements within their countries while IRG donors and international organization representatives were asked how important they considered these elements of the evaluation process.

The results are reported by *rank order of items rated very useful at the country-level.*

	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)
	Very useful	Very important	Somewhat useful	Somewhat important
1. Independence of the evaluation team	69%	77%	20%	14%
2. Composition and qualifications of the evaluation team	59%	73%	32%	14%
3. Position and responsibilities of the National Coordinator as separate from conducting the actual evaluation	54%	73%	32%	14%
4. The way in which the evaluation called attention to the aid effectiveness agenda of the Paris Declaration in participating countries	50%	68%	39%	23%

	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)
	Very useful	Very important	Somewhat useful	Somewhat important
5. Composition of the National Reference Group	46%	73%	31%	14%
6. Support from the international core evaluation team	48%	44%	73%	14%
7. Relationship between the National Reference Group and the evaluation team	45%	59%	36%	23%
8. Standardized matrix for evaluation data collection and reporting	45%	82%	42%	14%
9. Support from the Paris Declaration Evaluation Secretariat	44%	59%	30%	36%
10. The process for selecting the evaluation team	42%	82%	37%	4%

SUMMARY and OVERVIEW of OPINION ITEMS

Please respond with the country in mind. Keep in mind that your responses will be confidential. (n=80)		Strongly agree (%)	Agree (%)	Disagree (%)	Strongly disagree (%)	No opinion (%)
1.	Context is fundamental to understanding a country's findings.	81%	19%	-	-	-
2.	The contract process for the evaluation was consistent with the Paris Declaration principles.	30	51	4	5	10
3.	Country ownership of the evaluation was a priority.	43	46	9	-	2
4.	We received adequate support from the international core team when needed.	22	67	4	1	6
5.	The matrix provided a useful framework for collecting and reporting data.	26	62	8	2	1
6.	We felt that we could add country-specific items to the standard evaluation matrix.	23	64	4	1	8

- 100% strongly agreed or agreed that “Context is fundamental to understanding a country’s findings.”
- 89% strongly agreed or agreed that “Country ownership of the evaluation was a priority.”
- 88% strongly agreed or agreed that “The matrix provided a useful framework for collecting and reporting data.”
- 87% strongly agreed or agreed that “We felt that we could add country-specific items to the standard evaluation matrix.”

Ratings of the usefulness of the Standardized Operational Matrix

The survey of National Reference Group (NRG) members asked respondents to rate these various components in terms of their “usefulness to the country’s involvement in evaluation of the implementation of the Paris declaration.” Donors and representatives of international organizations participating in the International Reference Group (IRG) were asked in Copenhagen about the importance of these same elements.

	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)	Country-level NRG and evaluation team members (n=80)	Donors & international organization members (Copenhagen IRG, n=22)
	Very useful	Very important	Somewhat useful	Somewhat important
Standardized matrix for evaluation data collection and reporting	45%	82%	42%	14%

Dissemination at the Country Level

At the time of the survey, the country reports were three months old. Most respondents said that the dissemination process was just getting under way. Some expressed a high level of commitment to dissemination:

- “In my view, this [dissemination] is part and parcel of the whole exercise.”
- Others were dubious: “One of the weaknesses is that few stakeholders see the necessity for dissemination beyond the role of the study in the international evaluation.”

Here are the early data on dissemination efforts and plans at the country level.

Members of partner country national reference groups and evaluation teams (n=80)		Has happened more than once (%)	Has happened once so far (%)	Is planned but has not yet happened (%)	Is not planned at this point (%)	Not sure (%)
1.	Presentation to government officials	54%	15%	18%	4%	10%
2.	Presentation to donors	48	25	14	4	10
3.	Presentation to civil society organizations	26	24	32	6	11
4.	Public meeting to present and discuss the findings	16	19	29	18	19
5.	Press conference or other media communications and appearances (TV, radio, newspapers)	4	8	25	33	30
6.	Posted the report to the Internet	32	24	15	5	24

Understandably, at this early stage, the focus was on presentations to governments officials; 69% had made such presentations at least once. Wider dissemination efforts were still nascent for most. One respondent noted the challenge of getting beyond just national dissemination to regional dissemination with countries: “We all need to have much more dissemination activities on aid effectiveness at the provincial level in each country.”

Still, while relatively modest use had occurred at the time of the survey, respondents reported high expectations for eventual use.

Support from the Paris Declaration Evaluation Secretariat

National Reference Group members and evaluation team members were asked to rate the usefulness of the support received from the Paris Declaration Evaluation Secretariat to conduct the country evaluations (n=80). In rating the utility of Secretariat, it should be noted that the evaluation teams had little direct contact; their ratings would, at best, reflect a general sense they had picked up from others, like national coordinators.

The PD Evaluation Secretariat was useful to us in ...		Very useful to our evaluation process (%)	Somewhat useful to our evaluation process (%)	No support was needed in our evaluation process (%)	Don't know if it was needed or useful (%)
1.	understanding the purpose and importance of evaluating the implementation of the Paris Declaration.	59%	20%	12%	9%
2.	justifying (selling) and explaining the evaluation to national stakeholders.	39	35	15	11
3.	obtaining funding for this country's evaluation.	42	23	15	19
4.	disseminating the results of our evaluation.	34	30	10	25

Partner country responses, both national coordination and evaluation teams (n=80)	High capacity to engage in evaluation	Moderate capacity to engage in evaluation	Low capacity to engage in evaluation	Not sure about capacity
Paris Declaration Evaluation Secretariat	63%	19%	-	18%

COMPARISON HIGH CAPACITY RATINGS

Partner country responses comparing national coordinators and evaluation team leaders on capacity ratings	National Coordinators n=15	Evaluation Team Leaders n=15	Difference
HIGH CAPACITY TO ENGAGE IN EVALUATION			
Paris Declaration Evaluation Secretariat	87%	54%	33%

Ratings of Capacity to Engage in evaluation by Donors and International organization representatives in the IRG (n=23)

Donors and International organization representatives in the IRG (n=23)	Excellent	Good	Fair	Poor
Evaluation Secretariat.	70%	26%	-	-
Evaluation Management Group.	35	30	9	4
Core Evaluation Team.	52	39	9	-
Contributions of the International Reference Group.	13	52	30	-

Donors and International organization representatives in the IRG ratings of performance (n=23)

Overall, how would you rate performance of the	Excellent	Good	Fair	Poor
Evaluation Secretariat.	70%	26%	-	-
Evaluation Management Group.	35	30	9	4
Core Evaluation Team.	52	39	9	-
Contributions of the International Reference Group.	13	52	30	-

END of APPENDICES and END of REPORT